

Un modèle de choc
pour mesurer l'interaction entre événements ou entre biographies
Frédérique Letué (Laboratoire Jean Kuntzman, Université de Grenoble 2) et Valérie Golaz (INED)

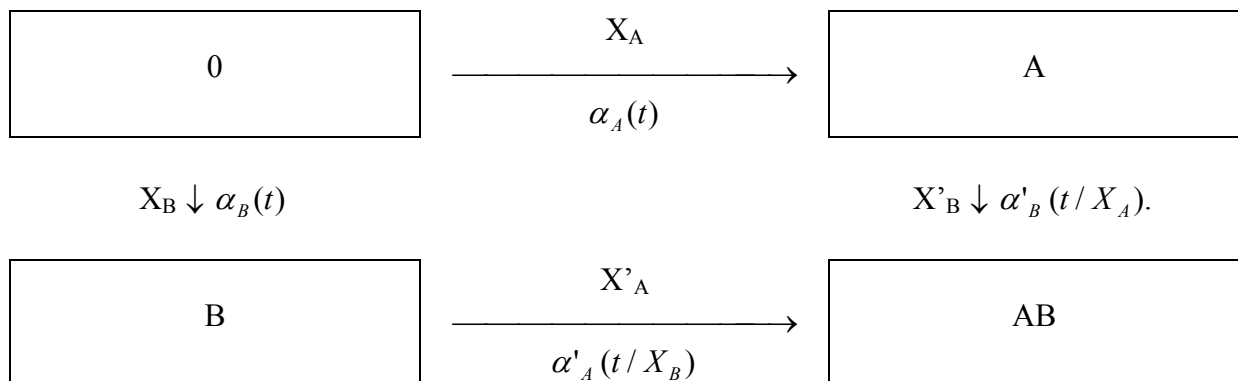
Ce papier présente un modèle semi paramétrique pour des couples de variables aléatoires en présence de censure. Ce modèle est appliqué à quelques exemples proches de l'analyse des biographies. Le même modèle peut être adapté et testé à la fois sur des couples d'événements non indépendants, au sein d'une trajectoire individuelle, et sur des événements survenant en parallèle sur des trajectoires individuelles liées. L'objectif est de mesurer la liaison entre trajectoires (par exemple trajectoires familiale et professionnelle d'un individu ou trajectoires professionnelles des deux membres d'un couple).

1. L'analyse des interactions entre événements : un rappel de la méthode usuelle avec l'exemple de Magenche

La concurrence entre deux événements est abordée dans Courgeau et Lelièvre (1989). Un individu part d'un état 0 où il n'a encore subi aucun événement (dans l'exemple développé ci après, l'individu est célibataire, non propriétaire). A partir de cet état, il peut rejoindre un état A (en union, non propriétaire) à une date X_A (date de l'entrée en union/ ou âge à l'entrée en union) ou un état B (célibataire, propriétaire) à une date X_B (date de l'accession à la terre ou âge à l'accession à la terre). Puis, à partir de chacun de ces états, il peut rejoindre un état final AB (en union, propriétaire), après une durée X'_B s'il vient de l'état A (durée entre le mariage et l'accession à la terre) ou après une durée X'_A s'il vient de l'état B (durée entre l'accession à la terre et le mariage).

Introduisons les notations caractérisant les lois des variables X_A, X_B, X'_A, X'_B :

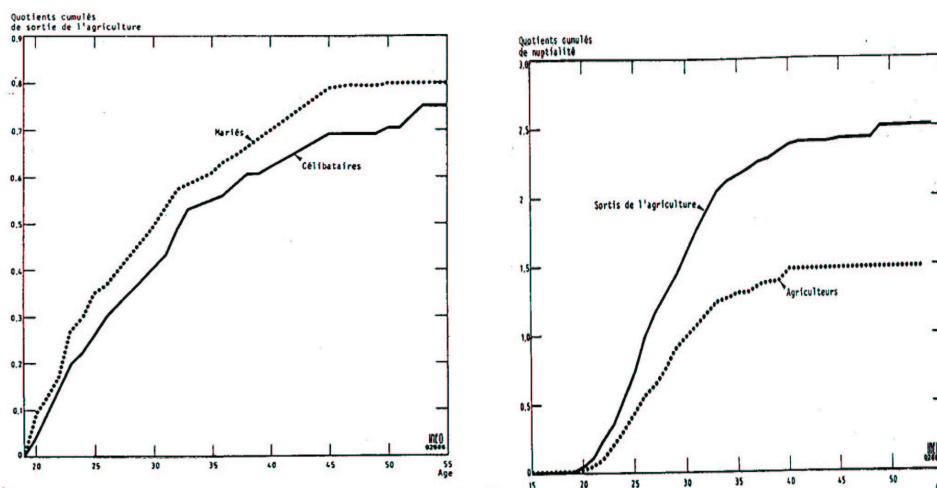
- La loi de X_A est caractérisée par son risque instantané α_A .
- La loi de X_B est caractérisée par son risque α_B .
- La loi de X'_A sachant X_B est caractérisée par son risque $\alpha'_A(t/X_B)$.
- La loi de X'_B sachant X_A est caractérisée par son risque $\alpha'_B(t/X_A)$.



où t est l'instant considéré, exprimé en durée écoulée depuis l'origine.

A et B sont indépendants si $\alpha'_A(t/X_B) = \alpha_A(t)$ et $\alpha'_B(t/X_A) = \alpha_B(t)$ (Courgeau et Lelièvre, 1989, p.76). En revanche, si l'une des égalités est fautive, il y a dépendance.

La mise en œuvre pratique proposée par les auteurs repose sur une méthode graphique : la comparaison des pentes des courbes de quotients cumulés. Individuellement, lorsque les données suivent l'hypothèse des risques proportionnels, chaque sens de l'interaction peut être évalué par un modèle de Cox prenant l'événement intermédiaire comme variable dépendant du temps.

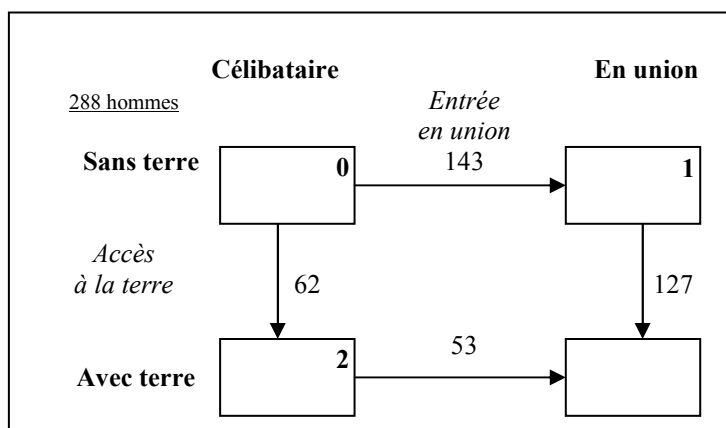


Ainsi, dans le cas des hommes, la comparaison des pentes des courbes de quotients cumulés de sortie de l'agriculture et de nuptialité montrent qu'il y a dépendance unilatérale : le mariage dépend du fait d'être sorti de l'agriculture, alors que la sortie de l'agriculture ne dépend pas du mariage (Courgeau et Lelièvre, 1986, p.311-312). Les auteurs évaluent le coefficient multiplicateur du risque de se marier, entre les agriculteurs et les hommes sortis de l'agriculture, selon le statut dans l'agriculture, entre 1,5 et 2,5.

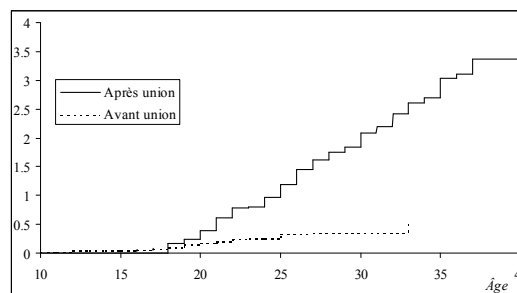
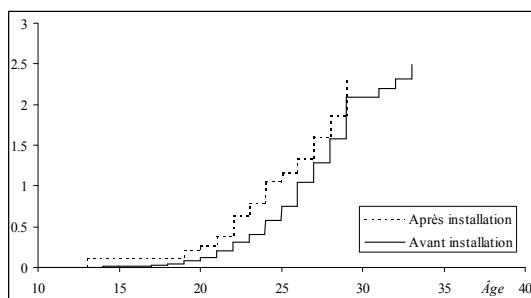
Cette méthode a aussi été validée sur de petits échantillons. Par exemple, dans l'enquête MaM¹, une relation de dépendance avérée est celle qui lie l'accès à la terre (le moment où un homme devient propriétaire par héritage ou allocation d'une partie des terres paternelles) et l'entrée en union (le moment où une femme s'installe chez son conjoint). Sur 288 hommes étudiés, du fait de la dispersion de l'échantillon en âge, 83 n'ont encore connu aucun de ces événements (universels) au moment de l'enquête.

¹ L'enquête Mobilité à Magenche comprend une partie biographique. Elle est présentée de manière complète sur http://www.ined.fr/fr/ressources_documentation/enquetes/liste_enquetes/bdd/enquete/mam_2008/

Entrée en union et accès à la terre



La comparaison des pentes des courbes de risques cumulés obtenues montre une relation de dépendance unilatérale dès 18-20 ans: l'accès à la terre n'a pas d'influence sur l'entrée en union, mais l'entrée en union multiplie le risque d'avoir accès à la terre.



Un modèle de Cox de l'accès à la terre, prenant le fait d'être en union en compte, peut être construit, mais nécessite le découpage de la période d'observation en deux, du fait du non respect des conditions de mise en œuvre de modèles à risques proportionnels sur la totalité de la période observée. Ainsi, les analyses doivent être menées séparément avant 25 ans et après cet âge, pour plusieurs co-variables pertinentes. Le fait d'être en union n'est pas significatif avant 25 ans, il le devient à partir de 25 ans : les hommes de plus de 25 ans en union ont un risque de devenir propriétaires 2,4 fois supérieur aux célibataires.

Cet exemple montre, par un échantillon très limité, à la fois l'intérêt et les limites de la comparaison visuelle, en particulier à la borne inférieure du graphique. La courbe ne tient pas compte de la faiblesse de l'échantillon en début de parcours (peu d'hommes de moins de 20 ans qui sont observés en union), et indique donc une tendance à ces âges qui n'est pas validée statistiquement. Si la comparaison visuelle demeure pratique, explicite, et peut servir de guide, il est utile de tester directement la significativité de la relation de dépendance entre les deux événements par une modélisation robuste de ce facteur de dépendance.

2. Un nouveau modèle pour mesurer les relations de dépendances

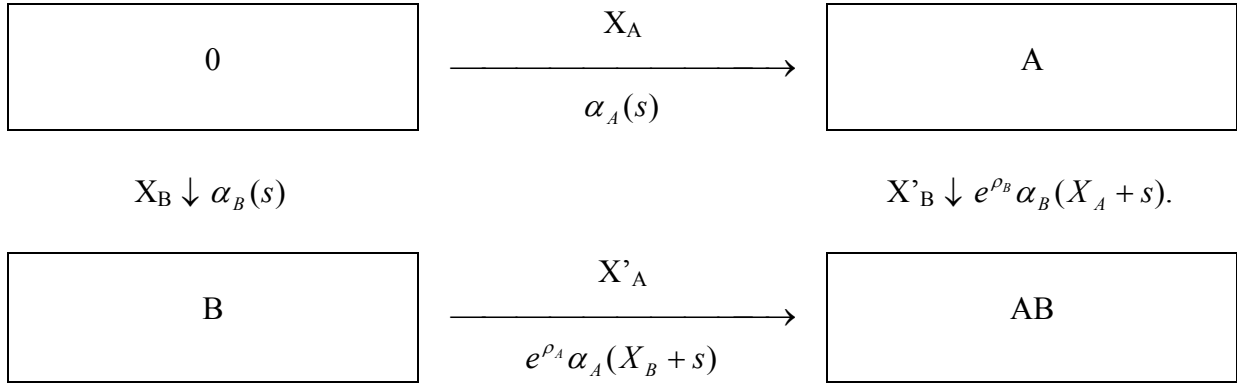
Alors que du point de vue pratique, la démarche précédente nécessitait la mise en œuvre de 7 étapes différentes (4 analyses non paramétriques, la représentation simultanée des courbes obtenues deux à deux, 2 modèles de Cox avec variable dépendant du temps), il est possible d'obtenir une mesure de l'interaction en modélisant de manière simultanée les rapports de

risques de l'un et l'autre événement. Les dépendances réciproques peuvent être mesurées par un modèle unique, qui évalue les interactions simultanément, dans les deux sens. Le modèle que nous proposons est le suivant : nous faisons l'hypothèse que le taux de risque instantané de connaître un événement en ayant déjà vécu l'autre est proportionnel à celui de la connaître sans avoir vécu l'autre.

$$\alpha'_A(t / X_B) = e^{\rho_A} \alpha_A(t) \text{ avec } t > X_B$$

$$\alpha'_B(t / X_A) = e^{\rho_B} \alpha_B(t) \text{ avec } t > X_A$$

Les coefficients ρ_A et ρ_B mesurent à quel point l'intervention de l'autre événement modifie le taux de risque de chacun de événement par rapport à s'il était seul.



Un individu donné passe soit par l'état A, soit par l'état B. On n'observe donc pas toutes les variables citées ci-dessus, mais suivant les cas :

- X_A et $X_A + X'_B$ s'il est passé par A
- X_B et $X_B + X'_A$ s'il est passé par B

Pour un individu donné, pour lequel on aurait une trajectoire non censurée, on peut dire qu'on observe

- $\tilde{X}_A = \min(X_A, X_B) + X'_A 1_{X_B < X_A}$
- $\tilde{X}_B = \min(X_A, X_B) + X'_B 1_{X_A < X_B}$

De plus, ces données peuvent elles-mêmes être censurées par des variables U_A et U_B (par exemple, par le décès ou le départ de l'échantillon d'observation ou la fin de l'étude) si bien qu'on observe finalement

- $T_A = \min(\tilde{X}_A, U_A)$
- $\delta_A = 1_{\tilde{X}_A < U_A}$
- $T_B = \min(\tilde{X}_B, U_B)$
- $\delta_B = 1_{\tilde{X}_B < U_B}$

A partir d'un échantillon de n observations $(T_{Ai}, T_{Bi}, \delta_{A,i}, \delta_{B,i})_{1 \leq i \leq n}$, on peut estimer les coefficients (ρ_A, ρ_B) qui donnent une mesure de l'interaction en maximisant le critère de log-vraisemblance partielle suivant, sur le modèle de la log-vraisemblance partielle de Cox :

$$l_n(\rho_A, \rho_B) = \sum_{h=A,B} \sum_{i=1}^n (\rho_h W_{h,i}(T_{h,i}) - \log S_{n,h}(\rho_h, T_{h,i})) N_{h,i}(\tau)$$

avec les notations :

- $W_{A,i}(s) = 1_{T_{B,i} < s}, W_{B,i}(s) = 1_{T_{A,i} < s}$,

- $S_{n,h}(\rho_h, s) = n^{-1} \sum_{i=1}^n e^{\rho_{h,i} W_{h,i}(s)} Y_{h,i}(s),$
- $Y_{A,i}(s) = 1_{T_{A,i} \geq s} (\delta_{B,i} 1_{T_{B,i} < s} + 1_{T_{B,i} \geq s}), Y_{B,i}(s) = 1_{T_{B,i} \geq s} (\delta_{A,i} 1_{T_{A,i} < s} + 1_{T_{A,i} \geq s})$
- $N_{A,i}(s) = \delta_{A,i} 1_{T_{A,i} \leq s} (\delta_{B,i} 1_{T_{B,i} < T_{A,i}} + 1_{T_{B,i} \geq T_{A,i}}), N_{B,i}(s) = \delta_{B,i} 1_{T_{B,i} \leq s} (\delta_{A,i} 1_{T_{A,i} < T_{B,i}} + 1_{T_{A,i} \geq T_{B,i}}),$
- τ étant l'instant final d'étude.

On remarque en particulier que

- un individu tel que $T_{A,i} < T_{B,i}$ contribue à la vraisemblance via la quantité $(\rho_B - \log S_{n,A}(\rho_A, T_{A,i})) \delta_{A,i} 1_{T_{A,i} \leq \tau} - \log S_{n,B}(\rho_B, T_{B,i}) \delta_{B,i} 1_{T_{B,i} \leq \tau} \delta_{A,i}$
- un individu tel que $T_{B,i} < T_{A,i}$ contribue à la vraisemblance via la quantité $-\log S_{n,A}(\rho_A, T_{A,i}) \delta_{A,i} 1_{T_{A,i} \leq \tau} \delta_{B,i} + (\rho_A - \log S_{n,B}(\rho_B, T_{B,i})) \delta_{B,i} 1_{T_{B,i} \leq \tau}$
- un individu tel que $T_{A,i} = T_{B,i}$ contribue à la vraisemblance via la quantité $-\log S_{n,A}(\rho_A, T_{A,i}) \delta_{A,i} 1_{T_{A,i} \leq \tau} - \log S_{n,B}(\rho_B, T_{B,i}) \delta_{B,i} 1_{T_{B,i} \leq \tau}.$

Ce dernier cas d'égalité ne devrait pas arriver théoriquement, mais il arrive que les conditions de recueil des données ne permettent pas de distinguer les deux dates.

Les propriétés asymptotiques des estimateurs ainsi trouvés ont été étudiées dans Letué (2008) ainsi qu'un estimateur de la matrice asymptotique de variance-covariance et que des estimateurs des taux de risques cumulés (intégrales de $\alpha_A(s)$ et $\alpha_B(s)$). Il est en particulier possible de construire des intervalles de confiance, tests, p-valeurs pour ces grandeurs.

a. Accès à la terre et entrée en union des hommes de Magenche avec la base de données totale

<p>n=145 U entrée en union, T accession à la terre U : 49 censures (49 individus sur 145 sont restés célibataires) T : 63 censures (63 individus n'ont jamais accédé à la terre.) U<T : 55 cas U=T : 62 cas U>T : 28 cas rhoU_chap : -0.1099941 (0.2484674) p-valeur : 0.6579898 rhoT_chap : 0.3012187 (0.3040904) p-valeur : 0.3219022</p>

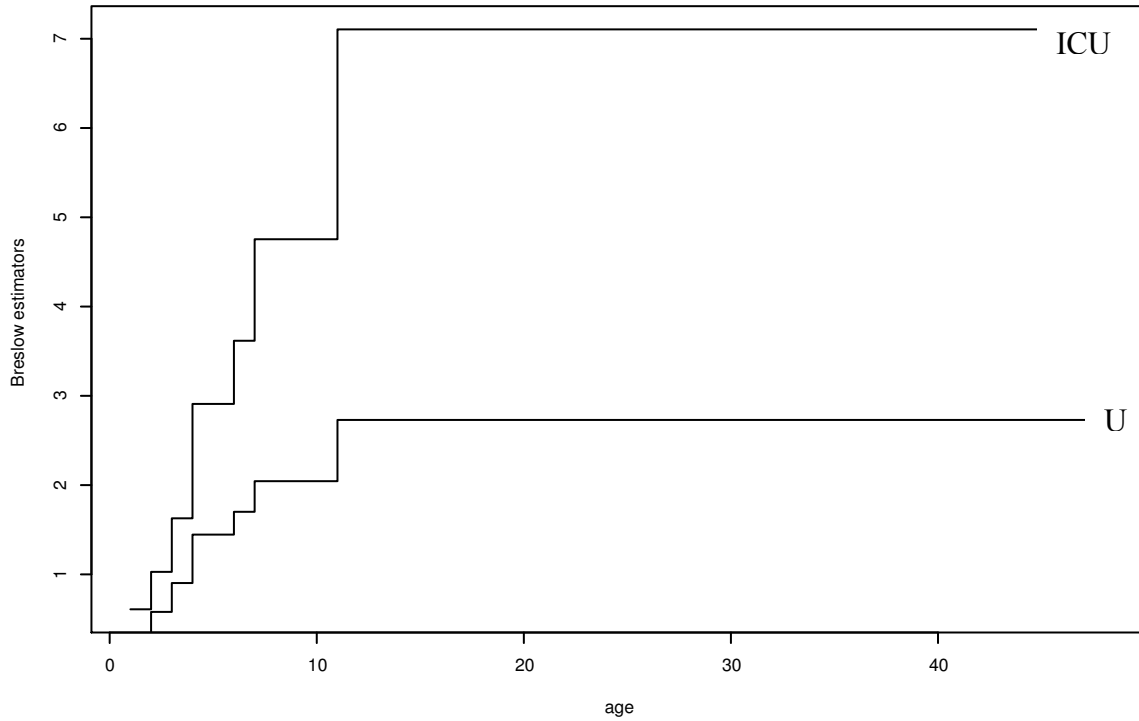
On trouve donc deux valeurs non significatives en travaillant sur la totalité de l'échantillon. En débutant à 15 ou 20 ans, les résultats sont similaires.

b. Accès à la terre et entrée en union des hommes de Magenche à partir de 25 ans

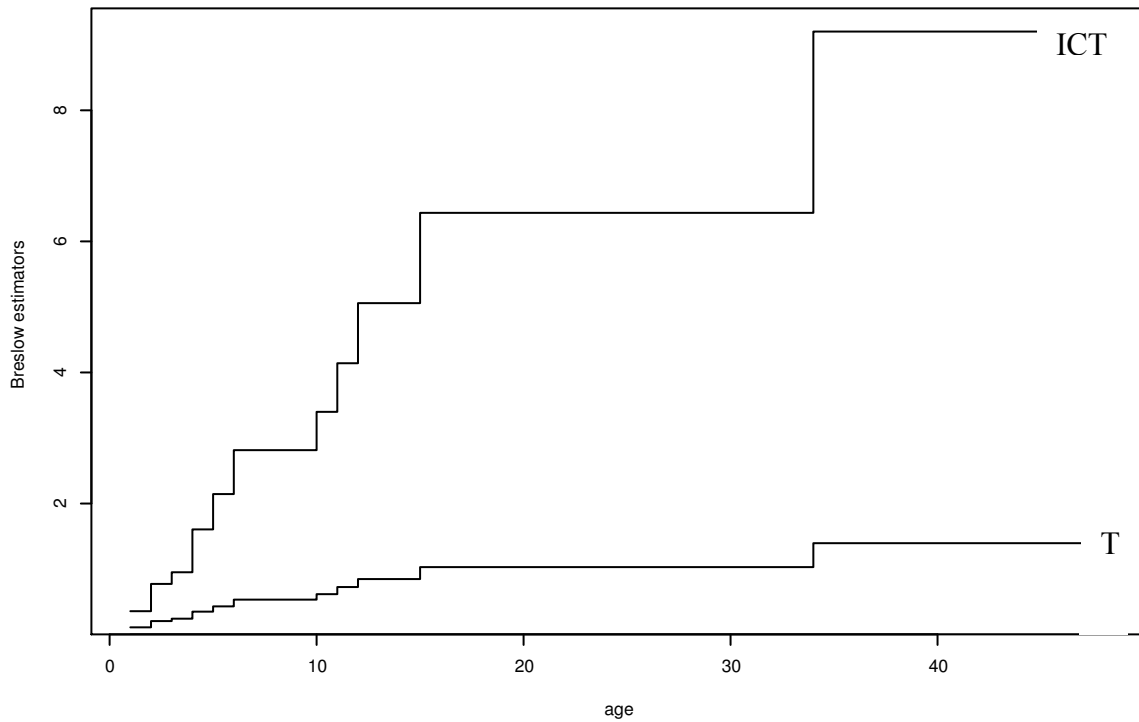
<p>n= 33 U : 5 censures T : 9 censures U<T : 19 cas U=T : 10 cas U : -0.7815904 (0.6336969) p-valeur 0.2174332 T : 1.0076981 (0.5784537) p-valeur 0.08149884</p>

Figure 4. Evaluation des risques hors interaction

Entrée en union :



Accès à la terre :



A 10 %, on retrouve donc les résultats obtenus par la méthode graphique : U non significatif, T significatif et positif, c'est-à-dire l'entrée en union a tendance à accélérer l'accès à la terre

des hommes qui n'ont pas encore connu ces événements à 25 ans. La méthode permet également de reconstruire le taux de risque cumulé de chacun des événements en l'absence de l'autre, avec leurs intervalles de confiance (figure 4)

3. Autres illustrations d'événements ou de trajectoires en interaction

Les exemples développés dans cette partie ont été tirés de l'enquête Biographie et entourage, réalisée en 2000-2001 sur un échantillon de 2830 personnes représentatif de la population d'île de France née entre 1930 et 1950. Cette enquête, ayant recueilli de manière systématique les trajectoires résidentielles, professionnelles et familiales des enquêtés, ainsi que des éléments de trajectoires des membres de leur entourage au fil de la vie, se prête particulièrement bien à l'étude des interactions entre événements au cours de la vie d'une personne, ou entre trajectoires individuelles de personnes proches l'une de l'autre (Lelievre et Vivier, 2001).

A. Evénements liés

L'interaction entre la naissance du premier enfant et le mariage

Pour les générations 1930-1950, enquêtées dans Biographies et entourage, le mariage demeure une étape importante : 88,3% des enquêtés ont été mariés au moins une fois au cours de leur vie passée. La question de l'influence de la naissance d'un enfant sur le mariage des parents, et du mariage sur la naissance d'un enfant ne semble pas en être une. La norme étant fortement en faveur de mariages précédant la constitution d'une famille, on s'attendrait spontanément à ce que le mariage favorise la fécondité. Pour l'influence d'une naissance sur un mariage éventuel, le résultat va moins de soi : un couple non marié ayant un enfant l'a peut-être fait sciemment, sans avoir l'intention de se marier pour autant, ce qui conduirait à une relation négative, mais dans d'autres cas, l'arrivée d'un enfant peut provoquer un mariage rapide, et alors la relation attendue est peut être positive.

L'analyse statistique confirme la première relation : à âge égal, le mariage accélère fortement l'arrivée du premier enfant. Elle tranche le débat sur la seconde : pour les femmes comme pour les hommes, une naissance hors mariage a un effet négatif sur l'éventualité d'un mariage... Il n'y a pas de différence significative pour l'estimation des coefficients rho entre hommes et femmes.

Echantillon total
n=2830
M : 332 censure (11,7%)
N : 317 censure (11,2%)
TM<TN : 1972 cas (78,5%)
TM=TN : 285 cas (11,3 %)
TM>TN : 256 cas (10,2%)
rho_N : 2.6465 (0.0547) p-valeur 0
rho_M : -0.4248 (0.0997) p-valeur 0.000020204

Hommes n= 1343 rho_N : 2.4979991 (0.0782944) p-valeur 0 rho_M : -0.4364766 (0.1425400) p-valeur 0.002197650
--

Femmes n=1487 rho_N : 2.649038 (0.07519036) p-valeur 0 rho_M : -0.407772 (0.1390409) p-valeur 0.003359767
--

Décohabitation de chez les parents et premier emploi
--

Les études portant sur le passage à l'âge adulte, tentant de définir un seuil de cette nouvelle phase de la vie, montrent que l'accès à un logement indépendant et à un premier emploi, de même que le début de la vie de couple, sont des éléments qui marquent des seuils généralement pris en compte. Le passage à l'âge adulte demeure cependant un seuil flou, difficile à établir de manière précise, et chacun de ces indicateurs généralement utilisés peut être remis en question, et forment eux-mêmes des seuils flous (Calvès et al., 2006). Ainsi entre le foyer parental et le logement indépendant, des allers retours réguliers et des situations d'hébergement temporaires sont souvent notés. Lorsqu'un jeune quitte temporairement le foyer parental, y revient, en repart, à partir de quand considérer qu'il est indépendant ? De même les emplois précaires qui se succèdent avant l'obtention d'un poste durable doivent-ils être pris en compte dans l'étude de l'accès à l'emploi ? Enfin, ces deux événements sont fortement liés, dans le sens où dans de nombreux cas, ils se succèdent rapidement dans le temps. Mais y a-t-il un lien réel entre premier emploi et décohabitation ? L'accès à l'emploi entraîne-t-il un départ de chez les parents plus précoce ? Le départ de chez les parents accélère-t-il la recherche d'un premier emploi ?

L'enquête Biographie et Entourage porte sur un échantillon de personnes nées dans les années 1930-1950. Les enquêtés ont donc pour la plupart connu ces événements dans la période de l'après guerre, une époque où globalement l'emploi était plus accessible et moins précaire qu'aujourd'hui. De même pour le logement. Les données se prêtent donc particulièrement bien à l'étude du troisième point : l'interaction entre accès au premier emploi et accès au premier logement indépendant.

1^{er} modèle. Avec tout l'échantillon, à partir de la naissance.

On travaille ici sur 2830 hommes et femmes, sachant qu'une centaine commence à travailler avant 14 ans, mais que relativement peu ont quitté le domicile parental (i.e. les deux parents) avant 12 ans. Se limiter à 12 ans, âge de la première activité professionnelle, ne change pas les résultats ci-dessous.

n=2830 E (Emploi): 47 censures (1,66%) L (Logement) : 634 censures (22 %)

TE # âge d'égo à son premier emploi censuré à l'âge d'ego au moment de l'enquête TL # âge d'égo à son premier logement censuré à l'âge d'ego au moment de l'enquête
--

TE<TL : 2022 cas (71,4%)

TE=TL : 373 cas (13,2%)

TE>TL : 435 cas (15,4%)

rhoE_chap : -0.2111072 (0.06229531) p-valeur : 0.0007019554

rhoL_chap : -0.3199339 (0.0527157) p-valeur : 1.286736e-09

Il y a beaucoup plus de personnes qui ont leur premier emploi avant de quitter le domicile parental que le contraire, ce qui correspond à nos attentes. Mais les deux coefficients sont significativement différents de 0 et, ce qui est relativement inattendu, ils sont négatifs. A âge équivalent, le fait d'avoir un emploi retarderait le départ, le fait d'avoir quitté le logement familial retarderait l'âge du premier emploi ... Une analyse différenciée hommes / femmes montre (ci-dessous), que ce phénomène est particulièrement fort chez les hommes : deux fois plus fort que chez les femmes en intensité, pour les deux sens de l'interaction.

Hommes

n=1343

E(Emploi) : 4 censures (0,3 %)

L(Logement) : 288 censures (21%)

TE<TL : 962 cas (71,6%)

TE=TL : 191 cas (14,2%)

TE>TL : 190 cas (14,2%)

rhoE_chap : -0.2681745 (0.09101122) p-valeur : 0.003212787

rhoL_chap : -0.4280793 (0.07642056) p-valeur : 2.123513e-08

Femmes

n=1487

E(Emploi) : 43 censures (2,9 %)

L(Logement) : 346 censures (23%)

TE<TL : 1060 cas (71,3%)

TE=TL : 182 cas (12,2%)

TE>TL : 245 cas (16,5%)

rhoE_chap : -0.1840370 (0.08575686) p-valeur : 0.03187044

rhoL_chap : -0.2177148 (0.07347944) p-valeur : 0.003047215

b. A partir de 16 ans

Ici on exclut les personnes qui se sont impliquées tôt dans la vie professionnelle, et les placements éventuels, en ne prenant en compte que les individus qui sont à 16 ans toujours au domicile parental et n'ont encore jamais travaillé.

n=2248

E(Emploi): 44 censures sur 2248 (1,9%)

L(Logement): 500 censures sur 2248 (22,2%)

TE<TL : 1506 (67%)

TE=TL : 326 (14,5%)

TE>TL : 416 (18,5%)

$\rho_E = -0.2111072$ (0.06229531) p-valeur 0.0007019554 $\rho_L = -0.3199339$ (0.05271570) p-valeur 1.286736e-09
--

On trouve à nouveau deux valeurs négatives, du même ordre que celles obtenues pour la population totale.

4. Aller plus loin : application du modèle à des parcours individuels en interaction

Deux événements dont l'interaction est discutée sont les décès dans le couple. Lorsqu'un homme meurt, l'espérance de vie de sa femme diminue-t-elle ? Lorsqu'une femme disparaît, le risque de décès de son conjoint augmente-t-il ? Les données de Biographies et entourage permettent de reprendre cette question, à partir des trajectoires des parents d'ego : le décès du père accélère-t-il celui de la mère et réciproquement ? Les analyses porteront sur les parents encore en couple aux 50 ans du père, et comporteront une variable explicative qui représente le décalage en âge entre homme et femme dans le couple. Une séparation conduit à la sortie d'observation (mais les cas de rupture sont rares).

D'autres pistes d'analyse de ce type d'interactions, entre trajectoires d'individus différents, peuvent être évoquées : changements professionnels de conjoints, à partir de l'entrée en union ; mobilité professionnelle (changement de fonction / de lieu de travail) au sein du couple, ... Un autre seuil flou apparaît cependant : celui de la trajectoire individuelle. Dès lors qu'un événement extérieur a une influence forte sur la trajectoire d'une personne, il perd son altérité et devient partie de la vie de la personne... D'où la difficulté d'identifier des interactions non individuelles, et de surcroît mesurables avec des données disponibles.

5. Aller encore plus loin avec des séquences d'événements en interaction

Le modèle pourra être adapté à des événements successifs (naissances successives / déménagements / emplois ...), la difficulté résidant alors dans l'écriture formelle du modèle et dans l'explosion potentielle du nombre de paramètres à estimer par rapport au nombre d'individus disponibles.

6. Conclusion

Le modèle présenté permet de mesurer par un couple de coefficients l'interaction entre deux éléments (événements d'une même trajectoire, de différentes trajectoires d'une même personne, ou de personnes différentes). Il peut tenir compte de variables supplémentaires (sexe par exemple). Il eut être utilisé pour des groupes d'interaction successives. L'une des difficultés majeures réside dans la collecte de données adaptées. Les enquêtes biographiques permettent l'étude d'événements en interaction au cours de la vie d'un individu. Mais dépasser le cadre de l'individu est une difficulté majeure, à la fois en raison de problèmes de définition (à partir de quand considère-t-on qu'un événement n'est pas dans la trajectoire de l'individu alors qu'il a eu une influence potentielle sur sa trajectoire ?) et de disponibilité de données. Des biographies de couples, ou de fratries, présentant une précision comparable dans les données collectées pour les personnes concernées, permettraient certainement d'apporter de l'eau à ce (nouveau) moulin.

Références

- Calvès A., Bozon M., Diagne A. et Kuepie M., 2006, Le passage à l'âge adulte : repenser la définition et l'analyse des premières fois, in GRAB, *Etats flous et trajectoires complexes. Observation, modélisation, interprétation*, Méthodes et Savoirs n°5, Ed. de l'INED/CEPED, Paris p.137-156.
- Courgeau D. et Lelièvre E., 1986, «Nuptialité et agriculture », *Population*, vol. 41, n° 2, p. 303-326.
- Courgeau D. et Lelièvre E., 1989, *Analyse démographique des biographies*, INED, Paris, 268p.
- Golaz, V., 2002, *Croissance démographique, pression foncière et diversification économique : une analyse biographique des stratégies de survie à Magenche (Gucha District, Kenya)*, thèse de doctorat de démographie économique, Institut d'Etudes politiques de Paris, 695p.
- Golaz V, C.Bonvalet, A.Diagne, D.Delaunay, F.Dureau, E. Lelièvre, 2006, « L'hébergement : un état sans statut », in GRAB, *Etats flous et trajectoires complexes. Observation, modélisation, interprétation*, Méthodes et Savoirs n°5, Ed. de l'INED/CEPED, Paris.
- GRAB (Groupe de recherche sur l'analyse des biographies), Antoine P. et E. Lelièvre (dir.), *Etats flous et trajectoires complexes. Observation, modélisation, interprétation*, Méthodes et Savoirs n°5, Ed. de l'INED/CEPED, Paris, 301p.
- Lelièvre Eva et Vivier Géraldine, 2001, « Evaluation d'une collecte à la croisée du quantitatif et du qualitatif : l'enquête Biographies et entourage », *Population*, vol.56, n° 6, p. 1043-1073.
- Letué, F., 2008, "A semiparametric shock model for a pair of event-related dependent censored failure times". *J. Statist. Plann. Inference* (2008), doi: 10.1016/j.jspi.2008.02.002