

**HOMOGENEISATION DE PROCESSUS OBSERVES :
Les programmes EPIDOR, EPICLA et EPIMAT.**

Laurence TRICOT
Unité d'Investigation Clinique
IUPG
6, rue du XXXI Décembre
CH - 1207 - GENEVE

Résumé :

Ces programmes servent à décrire un ensemble de processus observés, composés chacun d'un nombre variable d'événements survenus au cours du temps. On obtient différentes représentations des processus. L'une d'elles les caractérise par un nombre fixe de variables en vue de leur classification.

Mots-clés :

Processus, trajectoires, épisodes, événements, classification, temps, régularité.

1. Le programme EPIDOR

1.1. A quoi sert EPIDOR ?

EPIDOR a pour origine le traitement de données médicales : les processus de soins de patients psychiatriques, soins dispensés au cours du temps par un réseau de services, plus ou moins fréquents, plus ou moins rythmés. Ces processus sont souvent de longue durée, mais constitués d'événements peu nombreux et de divers types : par exemple, un patient consultera auprès de différents services une cinquantaine de fois pendant une période de 4 ans, durant laquelle il sera en plus hospitalisé une fois. En présence de telles données, les médecins s'interrogent sur le type de "trajectoires" suivies par leurs patients, autrement dit "l'histoire" de leur prise en charge.

Ainsi, EPIDOR a été utilisé pour la recherche de types de filières suivies par des patients ; en restant dans le domaine médical, ce programme a également servi à analyser la prise de médicaments au cours de l'hospitalisation.

D'une manière générale, EPIDOR peut être un instrument utile dans toute étude de prestations de services, dès que se posent les problèmes statistiques suivants :

- regroupement d'événements en "épisodes" au sein d'un processus,
- représentation de "trajectoires" ou processus observés,
- classification d'épisodes,
- classification de trajectoires.

EPIDOR permet d'analyser un ensemble d'individus, chacun étant caractérisé par une suite d'événements en nombre variable, survenus au cours du temps, suite que l'on appelle trajectoire. EPIDOR contient des sous-programmes utilitaires qui décrivent sous différentes formes les trajectoires après découpage de celles-ci en épisodes. De plus, ces sous-programmes préparent les fichiers nécessaires pour classifier les épisodes, puis les trajectoires, et pour effectuer des sorties graphiques sur plotter. Les trajectoires sont analysées les unes à la suite des autres. L'une des sorties d'EPIDOR est un fichier d'entrée d'un programme de classification de la bibliothèque Modulad : BOULES.

1.2. Quelques normes

Il s'agit de données liées au temps. Le temps est une variable qui prend des valeurs, ou dates, entières et non négatives, par exemple des secondes ou des jours.

Etant donné un individu i , on définit tout d'abord deux ensembles, et une application de l'un dans l'autre :

- un ensemble fini T_i de **dates** observées,
- un ensemble E d'états, ou **modalités**,
- une application x_i de T_i dans E .

L'ensemble des T_i sera inclus dans un intervalle de temps fixé par l'utilisateur : la **période d'observation**, qui peut, au choix, dépendre ou non de i .

Remarque : pour simplifier les notations, l'indice i n'apparaîtra pas dans ce qui suit lorsqu'il s'agira d'un individu quelconque.

1.2.1. Evénements, trajectoires, épisodes

- un **événement** : un couple $(t, x(t))$.
- un **événement de modalité A** : un couple (t, A) .
- une **trajectoire** : un ensemble d'événements
- une **trajectoire-A** : l'ensemble de tous les couples (t, A) d'une trajectoire.
- un **épisode** : une partie d'une trajectoire.
- un **épisode-A** : une partie d'une trajectoire-A.
- un **épisode continu** : un épisode dans lequel t prend toute valeur entière comprise entre deux entiers t_1 et t_2 . Le premier événement d'un épisode continu est une **entrée**, le dernier est une **sortie**.
- un **épisode discontinu** : s'oppose au précédent ; il peut y avoir des valeurs entières de t non consécutives pour les événements de cet épisode.
- un **épisode court** : épisode contenant un nombre d'événements inférieurs à un seuil donné.
- un **épisode long** : s'oppose au précédent. Seuls les épisodes longs seront traités par un programme de classification.

Exemple :

Remarque préliminaire : comme EPIDOR traite les individus successivement, il suffit pour avoir une idée des résultats de prendre l'exemple d'un individu unique.

Un patient est admis dans un hôpital le 12 octobre 1982 ; il en sort le 28 du même mois et, après ce séjour, consulte son médecin traitant pour des contrôles le 15 de chaque mois, depuis novembre 1982 jusqu'à mai 1983, puis une fois le 10 décembre 1983, et une dernière fois le 30 juin 1984. Il a été mis en arrêt de travail du 1er octobre au 15 novembre 1982. On lui a prescrit un médicament M1, à prendre quotidiennement durant le mois de

novembre, puis un médicament M2 à prendre une fois par semaine durant le mois de décembre à partir du premier.

Cette histoire se traduit ainsi, d'après les normes données plus haut :

a) l'unité de temps choisie est le jour de 24 heures.

b) la période d'observation sera par exemple du 1.1.82 au 31.12.83 (deux ans).

c) les événements :

- jour passé à l'hôpital le 12.10.82, jour passé à l'hôpital le 13.10.82 etc. (17 événements de ce type) ;

- consultation chez le médecin le 15.11.82, etc. (8 événements de ce type, la dernière étant hors période d'observation donc exclue) ;

- jour d'arrêt de travail le 1.10.82, etc. (46 événements de ce type) ;

- prise du médicament M1 le 1.11.82, etc. (30 événements de ce type)

- prise du médicament M2 le 1.12.82, le 8.12.82, le 15.12.82, le 22.12.82, le 29.12.82 ;

d) la trajectoire est l'ensemble de tous ces événements ;

e) les modalités : H : hôpital

C : consultation

A : arrêt de travail

M1 : médicament M1

M2 : médicament M2 ;

f) une trajectoire d'une modalité donnée : par exemple, la trajectoire-C est l'ensemble des 8 consultations datées.

g) épisodes continus : le séjour à l'hôpital, la période d'arrêt de travail, la prise du médicament M1 constituent des épisodes continus. Les dates du début (entrée) et de la fin (sortie) de l'épisode sont respectivement :

modalité	entrée	sortie
H	12.10.82	28.10.82
A	01.10.82	15.11.82
M	01.11.82	30.11.82

h) épisodes discontinus :

par exemple l'ensemble des consultations qui ont lieu le 15 de chaque mois jusqu'à mai 83 constitue un épisode discontinu, long. La consultation en décembre 83 constitue à elle seule un épisode discontinu court. Mais ceci n'est qu'un exemple : la définition de l'épisode dépend en effet de paramètres tels que l'intervalle de temps maximum entre deux événements consécutifs donnés par l'utilisateur.

1.2.2. Statut d'un événement

D'après le paragraphe 1.2.1., un événement est un couple "date-modalité" ; cependant, on introduit en fait une information supplémentaire pour chaque événement en ajoutant son **statut**. Cette variable nominale permet d'éviter, lorsqu'une trajectoire présente un épisode continu (voir 1.2.1.) de mettre dans les données tous les événements constituant cet épisode : on ne met que le premier, l'**entrée**, qui possède alors le statut 1, et le dernier, la **sortie**, qui possède le statut 2 ; le programme générera les événements intermédiaires. Les événements qui n'ont ni le statut 1 ni le statut 2, c'est-à-dire qui ne sont ni une entrée ni une sortie d'épisode continu, ont le statut 0.

Remarque : il peut arriver que les données contiennent une sortie sans qu'il y ait l'entrée correspondante parce que celle-ci est en dehors de la période d'observation des trajectoires. Le programme crée alors une **entrée fictive**, de statut 3, au début de la période d'observation. Parallèlement, si nécessaire, il crée une **sortie fictive**, de statut 4, à la fin de la période d'observation.

Résumé :

Statut	Evénement
1	entrée (donnée)
2	sortie (donnée)
3	entrée fictive (créée par EPIDOR)
4	sortie fictive (créée par EPIDOR)
0	autres cas (donnée)

Exemple :

dans l'histoire racontée ci-dessus :

- l'admission à l'hôpital, le premier jour d'arrêt de travail, le premier jour de prise du médicament M1 ont le statut 1 ;
- la sortie de l'hôpital, le dernier jour d'arrêt de travail, le dernier jour de prise du médicament M1 ont le statut 2.
- tous les autres événements ont le statut 0.

1.3. Fonctions des deux principaux sous-programmes

Les résultats sont obtenus en deux phases successives :

- a : le découpage en épisodes de chaque trajectoire,
- b : la mesure, pour chaque épisode, de différentes variables quantitatives, en vue d'une classification des épisodes.

Phase a : Découpage en épisodes

Le programme procède de la façon itérative pour chaque individu ; étant donné une modalité A, à chaque événement E de modalité A, il cherche une réponse à la question suivante : l'événement F de modalité A qui suit E dans l'ordre chronologique sur la trajectoire de l'individu considéré, peut-il appartenir à l'épisode-A contenant E, ou est-il le premier événement d'un nouvel épisode ? La réponse dépend des paramètres de commande, et du statut de E. Deux exemples :

- si les épisodes-A sont discontinus (information donnée dans les paramètres), E et F appartiendront au même épisode si leur distance dans le temps n'excède pas un seuil fixé (donné dans les paramètres) ;
- si les épisodes-A sont continus et si E a le statut 1 (entrée), F appartiendra au même épisode s'il a le statut 2, sauf si une règle de non interférence (voir 1.4) est entrée en jeu et oblige à tronquer l'épisode.

Phase b :

Mesure de variables quantitatives attribuées à un épisode : on effectue, selon le cas, tout ou partie de l'ensemble des mesures suivantes :

- nombre d'événements,
- durée totale,
- intervalle de temps **maximum** entre deux événements consécutifs,
- intervalle de temps **moyen** entre deux événements consécutifs,
- différence entre les deux mesures précédentes : indice de **régularité**,
- concentration des événements dans le temps : second indice de **régularité** (voir ci-dessous).
- **écart-type** des événements dans le temps.

En ce qui concerne la concentration, l'indice utilisé a été proposé initialement en vue d'analyser des données spatiales (voir bibliographie, 3^e référence). Il s'agit d'une série convergente (fonction EPACTE dans EPIDOR). Lorsqu'il y a moins de 20 événements, l'indice est approché par le premier terme de la série calculé comme suit : n événements ($n \leq 20$) étant représentés par des points sur l'axe des temps, on divise le plus petit segment contenant tous ces points en $n-2$ parties égales, ou mailles, et on compte le

nombre n_1 de mailles ne contenant aucun des $n-2$ points ; l'indice est le rapport $n_1/n-2$. L'indice est compris entre 0 et 1 ; plus il est proche de 0, plus la concentration est faible donc la régularité grande.

EPIDOR effectue les phases a et b, et fournit des sorties résumant les trajectoires ; il crée aussi des fichiers nécessaires à la classification des épisodes et des trajectoires.

1.4. Comment se présentent les données

EPIDOR a besoin, pour chaque événement, de trois informations :

- la date
- la modalité (voir 1.2.1)
- le statut (voir 1.2.2)

Par ailleurs, pour procéder au découpage en épisodes, un certain nombre de paramètres de commande sont nécessaires, dont voici les plus importants :

- le mode dans lequel sont données les dates : unité de temps ou jour-mois-an,
- l'origine des temps,
- les limites de la période d'observation,
- le nombre maximum d'événements d'une trajectoire,
- une éventuelle recodification des modalités introduites dans les données concernant les événements,
- pour chaque modalité recodée, son caractère continu ou discontinu (le caractère "mixte" est possible ; cela se produit lorsque, pour une même modalité, les événements peuvent avoir les statuts 0, 1, 2),
- en ce qui concerne les épisodes discontinus : leur durée minimum, le nombre minimum d'événements qu'ils contiennent, l'intervalle de temps maximum entre deux événements consécutifs d'un même épisode,
- le nombre minimum d'événements d'un épisode long (voir 1.2.1),
- des règles de non interférence (voir ci-dessous).

Les règles de non interférence sont des règles de construction des épisodes, dans le cas où l'on veut éviter l'interférence d'épisodes de modalités différentes : une trajectoire étant donnée, comportant des événements de modalités différentes, il arrive que l'on souhaite une description dans laquelle un épisode d'événements de modalité A, ou épisode-A, s'arrête à l'occurrence d'un événement de modalité B quels que soient les événements survenus par la suite.

Ce type de règle s'utilise si l'on veut, par exemple, éviter l'interférence d'épisodes continus et discontinus d'une même trajectoire. On indique donc pour chaque modalité concernée, l'identificateur A, suivi des identificateurs des modalités des événements dont l'apparition arrêtera l'épisode-A.

Exemple de données :

Reprenons l'exemple présenté au paragraphe 1.2.1., en ajoutant quelques données supplémentaires : *supposons qu'avant son arrêt de travail, le patient ait été consulter son médecin traitant le 20, le 24 et le 25 septembre 1982.*

Les données concernant le patient, outre l'identificateur de celui-ci, seront donc la liste des événements suivants, l'ordre chronologique n'ayant pas d'importance pour le programme :

Date	modalité	statut
20 09 82	C	0
24 09 82	C	0
25 09 82	C	0
12 10 82	H	1
28 10 82	H	2
15 11 82	C	0
15 12 82	C	0
15 01 83	C	0
15 02 83	C	0
15 03 83	C	0
15 04 83	C	0
15 05 83	C	0
10 12 83	C	0
30 06 84	C	0
01 10 82	A	1
15 11 82	A	2
01 11 82	M1	1
30 11 82	M1	2
01 12 82	M2	0
08 12 82	M2	0
15 12 82	M2	0
22 12 82	M2	0
29 12 82	M2	0

On suppose d'autre part les paramètres de commande suivants (valables pour tous les individus de l'étude) :

- mode dans lequel sont données les dates : jour - mois - an
- origine des temps : 01.01.82
- début et fin de la période d'observation : 01.01.82 et 31.12.83
- nombre maximum d'événements d'une trajectoire : 100
- pas de recodification de modalité

- caractère continu ou discontinu des épisodes de modalité donnée :
C, M2 : discontinu,
A, H, M1 : continu
- nombre minimum d'événements d'un épisode long : 6.
- règle de non interférence : un épisode-C (suite de consultations) s'arrête en cas d'admission à l'hôpital.
- seuils fixés pour les épisodes discontinus :

Modalité	durée minimum de l'épisode	nombre minimum d'événements	intervalle de maximum entre 2 événements
C	1	3	90
M2	1	1	15

Remarque : "la durée" d'un épisode est le nombre d'unités de temps qui séparent le dernier événement du premier, plus 1.

1.5. Exemple de résultat d'EPIDOR

a) A partir des données précédentes, EPIDOR donnera tout d'abord la liste des épisodes. Ici il y en a 6, par ordre chronologique du premier événement :

No de l'épisode	Modalité	caractère
1	C	discontinu, court
2	A	continu
3	H	continu
4	M1	continu
5	M2	discontinu, court
6	C	discontinu, long

Quelques remarques :

- l'épisode no 1 et l'épisode no 6 n'auraient formé qu'un seul épisode (les intervalles entre deux événements n'excédant pas 90 jours) s'il n'y avait pas eu la règle de non interférence, utilisée puisqu'il y a eu admission à l'hôpital.
- la consultation du 10.12.83 est trop éloignée du dernier événement de l'épisode no 6 pour faire partie de ce dernier ; elle aurait formé un épisode-C à elle seule s'il n'y avait pas eu le seuil de 3 événements au moins dans les paramètres.
- la consultation du 30.06.84 disparaît car elle est en dehors de la période d'observation.

b) En ce qui concerne l'unique épisode discontinu "long", EPIDOR calcule des variables quantitatives sur le vecteur des dates, converties en nombre de jours après l'origine des temps. Sur ce vecteur :

[319 349 380 411 439 470 500]

on obtient :

nombre d'événements : 7

durée totale : 182 j.

intervalle de temps maximum : 32 j.

intervalle de temps moyen : 31,17 j.

différence entre les deux intervalles : 0,83 j.

concentration dans le temps : 0

écart-type : 60,3 j.

Ces valeurs caractérisant l'épisode sont contenues dans les sorties listing et, après normalisation, feront partie des données du programme de classification.

c) En ce qui concerne les autres épisodes, les sorties listing contiendront la durée totale et le nombre d'événements.

2. Le programme EPICLA

C'est un petit programme qui prépare à partir des résultats d'EPIDOR les fichiers d'entrée d'un programme de classification de la bibliothèque MODULAD en vue de classer les **épisodes d'une modalité donnée**.

Les données sont dans un fichier de sortie d'EPIDOR ; il faut également introduire le nom de la modalité choisie.

Les sorties sont un fichier de données et un fichier de paramètres adaptés au programme BOULES de la bibliothèque MODULAD. Parmi les paramètres, il y a un rayon : EPICLA le calcule en divisant par 5 (nombre arbitrairement choisi) la plus grande distance entre un vecteur-ligne du tableau et le centre de gravité de tous les vecteurs-lignes.

Les données traitées par le programme de classification forment un tableau comportant autant de lignes que d'épisodes discontinus longs définis par EPIDOR à partir des trajectoires de tous les individus de l'étude. Les colonnes du tableau sont les variables quantitatives calculées sur les épisodes.

L'intérêt de cette interface est qu'il est facile à l'utilisateur de modifier EPICLA pour passer au programme de classification de son choix.

3. Le programme EPIMAT

EPIMAT est un petit programme qui utilise des résultats d'EPIDOR, d'EPICLA et un fichier de sortie standard de classification de MODULAD, pour caractériser les individus analysés par EPIDOR par un nombre fixe de variables : ceci afin de pouvoir classer les individus.

Sa fonction est de construire, pour une modalité donnée A, et pour un individu i, dont la trajectoire contient un événement de modalité A, un vecteur dont la dimension r est donnée dans les paramètres, valable pour tous les individus et toutes les modalités. Chaque composante de ce vecteur correspond à un intervalle de temps inclus dans la période d'observation. Ayant partagé celle-ci en r intervalles de temps égaux t_1, t_2, \dots, t_r , le vecteur sera construit ainsi :

- si la modalité A concerne les épisodes continus (voir 1.2.1) la k^e composante du vecteur sera le nombre d'unités de temps commune aux épisodes-A et à l'intervalle t_k , sur la trajectoire considérée.
- si la modalité A concerne des épisodes discontinus, la k^e composante sera un **numéro de classe** : la classe de l'épisode-A dont le premier événement appartient à t_k . Si plusieurs épisodes-A sont dans ce cas, seule la classe de l'épisode qui a le plus d'événements sera prise en compte. La classe est un résultat du programme de classification si l'épisode-A est long. Tous les épisodes-A courts sont mis ensemble dans une classe à part, qui a toujours le numéro 1.

Ainsi, pour la modalité A, l'individu sera caractérisé par r variables **quantitatives** si A est une modalité d'épisodes continus, et r variables **qualitatives** dans les autres cas.

Exemple :

Toujours avec l'ensemble traité plus haut, supposons $r = 8$, ce qui correspond à un partage des deux ans d'observation en trimestres.

Le vecteur à 8 composantes de la trajectoire-H sera :

[0 0 0 17 0 0 0 0]

4ème trimestre 1982

Le vecteur à 8 composantes de la trajectoire-C sera

$$[0 \ 0 \ 1 \ 2 \ 0 \ 0 \ 0 \ 0]$$

en notant 1 et 2 les numéros de classe des deux épisodes-C.

Ce vecteur à r composantes caractérisant un individu pour la modalité A est en lui-même une représentation de la trajectoire-A de l'individu. De plus, il constitue une ligne d'un tableau individus X variables, à traiter directement par un programme de classification en vue de classer cette fois les individus.

Le nombre r de colonnes de la matrice finale fait partie des paramètres de commande du programme EPIDOR.

Bibliographie

GIEL R., HERMANN P., TEN HORN G., TRICOT L.

Patterns of care and service utilization, chapitre V, Publication de l'OMS sur les services de santé mentale, European Series, Copenhagen, World Health Organization, 1987.

MODULAD, Logiciel d'analyse de données, brochure de documentation, INRIA, 1987.

TRICOT C., RAFFESTIN C.

Elaboration et construction d'un nouvel indice de concentration, l'Espace Géographique 1, no 4, 1974, pp. 303-310.

TRICOT L.

Evolution dans le temps du diagnostic et des soins en psychiatrie : étude statistique, Informatique et Epidémiologie Psychiatrique, ERES, 1985, pp. 233-243.

TRICOT L.

An example of the standardization of terminology : defining episodes of care. Psychiatric Case Registers in Public Health. Elsevier Science Publishers, 1986, pp 192-197.