

# EFFET CAUSAL EN PRÉSENCE DE LA MORT; APPROCHE PAR LE SYSTÈME STOCHASTIQUE

Daniel Commenges <sup>1</sup>

<sup>1</sup> *Equipe Biostatistique, INSERM U1219, ISPED 146 rue Léo Saignat, 33076 Bordeaux  
daniel.commenges@isped.u-bordeaux2.fr*

**Résumé.** Nous considérons le problème de l'évaluation de l'effet d'un facteur (qui peut dépendre du temps) sur un état physiologique qui peut être binaire (comme la démence) ou quantitatif (comme les capacités cognitives), dans des situations où le risque de mort n'est pas négligeable (comme dans les études sur le vieillissement). Notre approche est basée sur une approche dynamique de la causalité. Cette approche utilise le formalisme des processus stochastiques et le concept de système, et nous l'appelons "l'approche de la causalité par le système stochastique". Nous montrons que nous devons d'abord regarder l'effet du facteur sur le risque de mort et ensuite l'effet sur l'état physiologique en utilisant le concept de "dépendance locale" sur un horizon aléatoire défini par le temps de mort. Dans cette approche nous devons clairement distinguer le système et les observations. Nous donnons un exemple sur les capacités cognitives dans le vieillissement.

**Mots-clés.** causalité, épidémiologie, système stochastique

**Abstract.** We aim to investigate the issue of estimating the effect of a factor (which may be time-dependent) on a physiological state which can be binary (such as dementia) or quantitative (such as cognitive ability) in situations where the death risk is not negligible (such as in ageing studies). Our approach is based on the dynamic approach to causality. This approach uses the formalism of stochastic processes and the concept of system, and we call it "the stochastic system approach to causality". We show that we must first look at the effect on death, and then we look at the effect on the physiological state using the concept of "local dependence" on a random horizon defined by the time of death. In this approach we clearly distinguish between the system and the observations. An example is given about cognitive ability in the elderly.

**Keywords.** causality, epidemiology, stochastic system

## 1 L'approche de la causalité par le système stochastique

Les statisticiens s'intéressent beaucoup depuis quelques années à la formalisation de la causalité, avec des applications en particulier à l'épidémiologie et à l'économie. L'approche

contrefactuelle a été beaucoup développée. Une autre approche, davantage dans la continuité de la modélisation en physique (et aussi reliée à la causalité selon Granger) est basée sur des systèmes dynamiques. Elle a été développée principalement par Aalen (1987), Didelez (2008), Commenges et Gégout-Petit (2009) et Commenges et Jacqmin-Gadda (2015). Les principales idées de cette approche sont

- représenter les états physiologiques par des processus stochastiques
- formaliser l’influence d’un processus sur un autre par le concept d’influence locale
- pour que ces influences puissent être considérées comme causales, il faut que le système soit suffisamment riche
- clairement distinguer entre le système, représenté par l’ensemble des processus stochastiques (ou un processus multivarié) et les observations que l’on en fait

Donnons un exemple concernant les capacités cognitives et l’hypertension chez les personnes âgées. Ces deux états physiologiques peuvent être représentés par des processus stochastiques,  $Y$  et  $V$  respectivement. Supposons que pour un certain sujet ces processus évoluent suivant une loi qui peut être caractérisée par leur décomposition de Doob-Meyer sous forme différentielle:

$$\begin{aligned} dV_t &= \beta_1 dt + \omega dB_t \\ dY_t &= I_{Y_t=0} \alpha_0(t) e^{\gamma V_t} dt + dM_t, \end{aligned}$$

où  $(M_t)$  et  $(B_t)$  sont des martingales. Le concept de dépendance locale (WCLI pour “weak conditional local independence”) est basé sur la mesurabilité de la partie prévisible de la décomposition, que nous appelons “compensateur”; nous appelons “intensité” sa différentielle. Dans cet exemple nous voyons que  $V_t$  apparait dans l’intensité de  $Y$ , donc celle-ci ne serait pas mesurable dans une filtration ne contenant pas  $V$ . Nous pouvons dire que  $Y$  n’est pas localement indépendant de  $V$ , ou de manière équivalente, que  $V$  influence  $Y$ . Au contraire, on voit que  $Y$  n’apparait pas dans l’intensité de  $V$ . On dit que  $V$  est localement indépendant de  $Y$ , ou que  $Y$  n’influence pas  $V$ , et l’on note:  $V \rightarrow_{\mathbf{X}} Y$  et  $Y \not\rightarrow_{\mathbf{X}} V$ . Un graphe peut être tracé en mettant une flèche pour représenter les influences. Le graphe n’est pas un DAG (directed acyclic graph) car il n’est pas nécessairement acyclique.

La dépendance locale peut-être étudiée sur un horizon fini, déterministe ou aléatoire.

## 2 Le rôle de la mort dans un système

Dans les études de vieillissement le risque de mort ne peut être négligé; c’est aussi le cas lorsque l’on étudie des maladie graves comme le cancer et le concept de mort peut-être étendu à des systèmes complexes autres que des sujets humains. L’idée clé, et d’ailleurs

évidente, est que les processus décrivant l'état du système ne sont *définis* que si le système existe, c'est à dire, si le sujet n'est pas mort. Le statut vital peut être représenté par un processus de dénombrement  $D$  qui prend les valeurs 0 si le sujet est en vie, 1 sinon. Tous les autres processus sont définis sur la période  $0, T_D$ , ou  $T_D$  est le temps de mort. Dans notre exemple, nous pouvons ajouter le processus  $D$  et nous intéresser aux indépendances locales (ou aux influences) à l'horizon  $T_D$ . Donc il n'y a pas symétrie entre le processus  $D$  et les autres processus.

La conséquence sur le plan de l'interprétation est que si nous pouvons manipuler le processus  $V$ , nous devons d'abord regarder son influence sur le processus vital  $D$ , et ensuite son influence sur le processus  $Y$  à l'horizon  $T_D$ .

### 3 Illustration

Prenons l'exemple de l'évolution des performances cognitives avec le vieillissement, et en particulier l'effet d'un facteur  $V$  sur ces performances;  $V$  peut représenter l'hypertension, qui peut être un processus à espace d'état binaire ou quantitatif. Pour pouvoir interpréter de manière causale l'influence de  $V$  sur  $Y$ , nous devons considérer un système assez riche  $\mathbf{X} = (Y, V, C, G', G'', D)$ ; essentiellement la condition est qu'il n'y ait pas de facteur de confusion, c'est à dire que l'on ne peut pas construire un système plus grand dans lequel un facteur  $U$  influencerait à la fois  $Y$  et  $V$ . Il faut donc inclure dans le système des processus  $C$  et des facteurs constant, appelés "attributs"  $G'$  qui peuvent influencer  $Y$  et  $V$ ; il est souhaitable d'inclure des facteurs  $G''$  qui influencent uniquement  $Y$ , même si de tels facteurs ne sont pas observés, ils peuvent être représentés par des effets aléatoires;  $G'$  et  $G''$  peuvent en particulier être des facteurs génétiques (incluant le sexe).

Nous pouvons comme Ganiayre, Commenges and Letenneur (2008) considérer que les performances cognitives représentées par  $Y$  sont observées de manière indirecte par un ou plusieurs tests psychométriques et également par un diagnostic de démence. Les scores aux tests psychométriques sont relevés à des temps discrets, à l'occasion de visites planifiées dans une cohorte (la cohorte Paquid pour le travail cité). Le lien entre  $Y$  et les observations est basé sur une catégorisation; l'observation fournie par le test  $k$  au temps  $t_j$  est:

$$Z_{t_j}^k = q \quad \text{si et seulement si} \quad c_q^k \leq Y_{t_j} + \varepsilon_j^k < c_{q+1}^k, \quad (1)$$

avec  $c_0^k = -\infty$  et  $c_Q^k = +\infty$ . Pour le diagnostic de démence il s'agit d'une catégorisation binaire. Les paramètres  $c_q^k$  doivent être estimés en plus des paramètres du modèle de la dynamique du système. Ganiayre, Commenges and Letenneur (2008) n'ont pas tenu compte de la mort. Des modèles peuvent être développés pour tenir compte à la fois de l'observation de tests psychométriques et de la mort comme dans Rouanet et Coll. (2016).

## Bibliographie

- [1] Aalen (1987), *Dynamic modelling and causality*, Scandinavian Actuarial Journal, 1987, 177-190.
- [2] Didelez, V. (2008), *Graphical models for marked point processes based on local independence*, Journal of the Royal Statistical Society: Series B, 70, 245–264.
- [3] Commenges, D. et Gégout-Petit, A. (2009), *A general dynamical model with causal interpretation*, Journal of the Royal Statistical Society: Series B, 71, 719-736.
- [4] Commenges, D. et Hélène Jacqmin-Gadda (2015), *Dynamical Biostatistical Models*, Chapman & Hall.
- [5] Ganiayre, J., Commenges, D. and Letenneur, L. (2008), *A latent process model for dementia and psychometric tests*, Lifetime Data Analysis, 14, 115–133.
- [6] Rouanet, A., Joly, P., Dartigues, J-F., Poust-Lima, C. and Jacqmin-Gadda, H. (2016) *Joint latent class model for longitudinal and interval-censored semi-competing events: Application to dementia multistate data*, soumis.