

Deux modèles de Markov caché pour processus multiples et leur contribution à l'élaboration d'une notion de style postural

Antoine Chambaz¹, Isabelle Bonan² et Pierre-Paul Vidal³

Title

Two hidden Markov models for multiple processes and their contribution to the definition of a postural style

Résumé

Nous décrivons dans cet article deux modèles de Markov caché (HMMS) pour processus multiples que nous mettons au point à des fins d'application à un problème biomédical d'étude du maintien postural. Informellement, ces modèles sont des HMMS standard considérant simultanément plusieurs processus individuels. Ceux-ci partagent les mêmes distributions de chaîne de Markov cachée et d'émission, avec l'ajout éventuel d'un effet de type mixte dans la partie conditionnelle du modèle. Nous discutons des aspects théoriques de l'estimation de l'ordre des deux modèles, ainsi que des aspects numériques. Nous contribuons enfin à l'élaboration d'une notion de "style postural".

Mots-clés : BIC, estimation de l'ordre, modèle de Markov caché (HMM), posture

Abstract

Two hidden Markov models (HMMS) for multiple processes are introduced and applied to the biomedical study of how people maintain posture. Informally, they model simultaneously multiple processes by making them share the same hidden Markov and observed conditional distributions. One of them incorporates covariates and random effects in the conditional part. We discuss the theoretical issue of order estimation for both models as well as computational issues. Finally, we contribute to the definition of a notion of "postural style".

Keywords : BIC, hidden Markov models (HMM), order estimation, posture

Mathematics Subject Classification: (62M02, 62M05, 62P10)

¹MAP5, Université Paris Descartes, Antoine.Chambaz@parisdescartes.fr

²LNRS, Université Paris Descartes, Hôpitaux Fernand-Widal et Lariboisière, Isabelle.Bonan@chu-rennes.fr

³LNRS, Université Paris Descartes, Pierre-Paul.Vidal@univ-paris5.fr

1 Introduction

Les modèles de Markov cachés (“Hidden Markov models” en anglais, d’où l’abréviation HMMs) sont un outil très apprécié pour la modélisation de données longitudinales. Selon ces modèles, un processus caché (c’est-à-dire non observé) $\{Z_t : t = 1, \dots, n\}$ coexiste avec le processus observé $\{Y_t : t = 1, \dots, n\}$. Le processus $\{Z_t\}$, supposé être une chaîne de Markov, convoie toute l’information relative à la dépendance : conditionnellement à $\{Z_t\}$, le processus observé est à coordonnées indépendantes — plus précisément, conditionnellement à Z_t , Y_t est indépendante de $\{Y_s, Z_s : s \neq t\}$. Ainsi, la distribution de $\{Y_t\}$ est caractérisée par la loi initiale et le noyau de transition de $\{Z_t\}$ d’une part, et par les lois conditionnelles de Y_t sachant Z_t d’autre part. Lorsque Z_t ne prend qu’un nombre fini de valeurs, la loi marginale de Y_t s’écrit comme un mélange fini. À ce titre, les HMMs offrent un moyen de modéliser la sur-dispersion. Évidemment, si $\{Z_t\}$ est une suite de variables indépendantes identiquement distribuées (i.i.d), alors $\{Y_t\}$ est elle-même une suite de variables i.i.d.

La littérature dédiée à l’étude des HMMs est foisonnante. Nous renvoyons le lecteur à la monographie [6] pour une présentation mathématique très complète du thème. Les exemples d’utilisation des HMMs sont nombreux, nous citerons parmi tant d’autres [21] pour une application à la reconnaissance vocale, [17] pour des applications à l’étude de données post-génomiques, [1] pour une application à l’étude du nombre de lésions chez un patient atteint de sclérose en plaques.

En général, un unique processus $\{Y_t\}$ (éventuellement multidimensionnel) est modélisé par un modèle de Markov caché. Il serait pourtant souvent pertinent de modéliser simultanément plusieurs processus $\{Y_t^i\}, i = 1, \dots, N$, dans l’esprit des modèles de Markov cachés. Partant de ce constat, Altman [2] a développé une extension des HMMs. Appelée HMMs mixtes, cette extension repose sur l’introduction de covariables et d’effets aléatoires dans les deux parties (la cachée et la conditionnelle) du modèle de Markov caché standard. Appliquée à l’étude du nombre de lésions chez *plusieurs* patients (et non plus un unique) atteints de sclérose en plaques, l’extension permet de bien mieux comprendre les mécanismes de la maladie et des différents traitements qui sont en jeu.

Notre travail, motivé par une application biomédicale, repose sur deux modèles de Markov caché pour plusieurs processus (d’où l’expression “HMM pour processus multiple”) qui sont des cas particuliers de HMM standard (caractérisé par une matrice de transition creuse) et de HMM mixte. L’application biomédicale en question est présentée dans la Section 2. Le premier modèle statistique sur lequel l’étude repose est décrit dans la Section 3 et nous exposons dans la Section 4 les résultats de son ajustement à nos données. Nous y argumentons l’intérêt d’introduire un second modèle, présenté dans la Section 5, sa contribution à l’élaboration d’une notion de style postural étant l’objet de la Section 6. Une discussion de notre approche numérique et les démonstrations de l’article sont finalement reportées en Annexe.

2 La recherche d'une notion de style postural : un enjeu important

2.1 Éléments biomédicaux

Le contrôle postural d'un individu est fondé sur trois types d'informations encodées par les systèmes visuel, vestibulaire (situé dans l'oreille interne) et proprioceptif (composé par les récepteurs sensoriels situés au voisinage des os, des articulations et des muscles et sensibles aux stimulations produites par les mouvements du corps). La façon dont le système nerveux central traite ces informations varie avec l'expérience sensorimotrice de chaque individu. Elle est modelée notamment par l'âge, les sports et les professions pratiqués, ainsi sans doute que par des facteurs génétiques. Un des éléments fondamentaux du contrôle postural tient à la capacité du cerveau central à saisir l'information sensorielle la plus pertinente à un instant donné pour réagir à une situation donnée.

Dans ce contexte, on peut comprendre que chacun tend, en raison de son expérience sensorimotrice, à développer une préférence pour un type d'information sensorielle particulier, qui se retrouve ainsi sollicitée de façon prédominante. La préférence visuelle est sans doute la plus fréquente, en tout cas la mieux décrite. On peut la relever chez des sujets sains mais elle est surtout courante chez les personnes âgées, chez les personnes souffrant de la maladie de Parkinson ou suite à un accident vasculaire cérébral ou à une pathologie vestibulaire. Or si une telle sélection systématique d'un mode perceptif permet à un individu de se déplacer efficacement dans son environnement habituel, il est clair qu'elle est peu adaptée pour répondre à des situations nouvelles ou inattendues, pour ne pas dire qu'elle est dangereuse. Un tel mode de fonctionnement est potentiellement plus susceptible d'entraîner une chute, chute dont les conséquences sont souvent dramatiques au-delà de soixante ans puisque les chutes et leurs séquelles entraînent le décès de près de dix mille personnes par an. Confronté à ce problème, le rééducateur tâche d'identifier, sur des bases cliniques et empiriques, une "préférence sensorielle" du patient qui lui a été adressée pour des troubles du contrôle postural. Il s'efforce ensuite de les corriger en amenant le sujet à prendre en compte la totalité de ses afférences sensorielles pour réguler sa posture. La "préférence sensorielle", un concept encore flou, n'est pas encore validée par des mesures quantitatives et son éventuelle modification par la rééducation non plus.

Ce travail a pour objet de contribuer à l'élaboration d'une notion de "style postural".

2.2 Un aperçu de l'état de l'art

L'étude du contrôle postural est l'objet d'un grand nombre de publications. On peut très schématiquement les répartir en deux catégories, selon qu'il y est question de la modélisation du processus d'intégration des informations (voir le récent [18] et ses références) ou bien de l'analyse statistique des déplacements d'un sujet (voir [26] pour une étude statistique par régression, pionnière du genre, de la *démarche* – et non de la *posture*). Notre approche s'inscrit dans ce second groupe.

Toujours schématiquement, les données ici analysées statistiquement sont obtenues sous la forme d'enregistrements des petits déplacements d'un individu se tenant debout campé sur ses

deux pieds et soumis à diverses perturbations. Chaque type de stimulation vise à explorer un des trois systèmes d'acquisition de l'information :

- la fermeture des yeux et la stimulation optocinétique permettent de s'intéresser au rôle du système visuel (voir par exemple [23] et ses références) ;
- les stimulations galvaniques permettent d'étudier le système vestibulaire (voir par exemple [13, 19] et leurs références) ;
- les stimulations vibratoires permettent de se pencher sur le rôle du système proprioceptif (voir par exemple [13, 15] et ses références).

La quantification du contrôle postural est obtenue via une plateforme de force. Cet appareil, semblable à un pèse-personne, évalue les positions successives du point de pression maximal exercé par le sujet.

C'est en abordant l'éventail des techniques statistiques développées pour étudier le contrôle postural qu'on peut mettre en exergue en quoi notre projet se démarque de ses prédécesseurs.

L'argumentation statistique d'un grand nombre d'articles est fondée sur des comparaisons de quantités moyennes ou extrémales (déplacement moyen, vitesse moyenne, étendue min-max etc.). Ces quantités ont le défaut de ne pas exploiter suffisamment la dynamique temporelle du phénomène.

Cette dynamique est en revanche centrale dans les articles qui exploitent les modèles classiques de séries temporelles (comme notamment dans [14, 28]). C'est à plus forte raison aussi le cas des travaux dans lesquels les trajectoires des positions du point maximal de pression sont modélisées par un mouvement brownien (voir [8] et ses références). Dans leurs prolongements, des modélisations par mouvement brownien fractionnaire introduites dans [9] ont aussi été exploitées ; les auteurs de [4, 3] étudient ainsi la façon dont la dépendance se propage à travers le temps au sein d'une trajectoire (pour des protocoles sans stimulation). Un autre courant met enfin en jeu des modélisations par diffusion, typiquement de type Ornstein-Uhlenbeck (voir en particulier [25, 12]).

Nous accordons aussi une place centrale à la dynamique temporelle du phénomène dans notre étude.

2.3 Protocoles médicaux d'acquisition de données

Nous n'exploitons dans ce travail qu'une partie des données produites par l'équipe médicale dont nous sommes partenaires pour cette étude.

Notre choix s'est porté sur l'un des protocoles qui ont été à ce jour appliqués à tous les participants à l'étude, qu'ils soient des sujets *sains* ou des patients *vestibulaires* ou *hémiplegiques*. Il entre dans la catégorie des protocoles à stimulations vibratoires.

Les patients se tiennent debout *les yeux fermés* sur une plateforme de force, appareil dont nous avons déjà écrit qu'il s'apparente à un pèse-personne. Cet appareil enregistre à intervalles réguliers les positions successives des points où le pied gauche et le pied droit du patient exercent séparément une pression maximale. La fréquence d'échantillonnage est de 40Hz (soit un relevé toute les $\delta = 25$ millisecondes).

Le protocole se découpe en trois phases expérimentales : lors des 15 premières secondes,

le maintien n'est pas perturbé artificiellement ; suivent alors 35 secondes durant lesquelles le maintien est perturbé par stimulation vibratoire des triceps du patient ; pendant les 20 dernières secondes, le maintien n'est finalement plus perturbé artificiellement.

2.4 Description succincte des données

Nous observons $N^{(0,0)} = 32$ patients sains, $N^{(1,0)} = 23$ patients vestibulaires et $N^{(0,1)} = 16$ patients hémiplésiques, soit un total de $N = 71$ patients.

Les observations brutes pour l'un d'entre eux s'écrivent $\{U_{t\delta}^G, U_{t\delta}^D\}$, $U_{t\delta}^G$ (respectivement $U_{t\delta}^D$) étant la position où le pied gauche (respectivement droit) exerce une pression maximale à l'instant $t\delta$. Ces positions sont repérées par leurs coordonnées cartésiennes.

Nous procédons à un premier résumé de ces données brutes en ne considérant que la suite des milieux des segments $\{U_{t\delta} = \frac{1}{2}(U_{t\delta}^G + U_{t\delta}^D)\}$. Heuristiquement, cette suite décrit la trajectoire de la projection du centre de gravité du patient sur la plateforme de force.

Pour avoir observé sur plusieurs trajectoires des sursauts en début de protocole (la plupart du temps lors de la première seconde), nous ne prenons pas en compte les deux premières secondes d'enregistrement.

Par ailleurs, une étude préliminaire a confirmé la pertinence de résumer ces données en ne considérant que la suite $\{X_{t\delta}\}$ des distances séparant $U_{t\delta}$ d'un point de référence. Possibilité parmi plusieurs, ce dernier est défini comme la valeur médiane des $U_{t\delta}$ sur la première période (c'est-à-dire avant les stimulations).

Cette même étude préliminaire nous a enfin permis de mettre en évidence que la volatilité de la suite $\{X_{t\delta}\}$ est particulièrement intéressante. Partant de ce constat, nous effectuons une dernière transformation en introduisant pour tout t la quantité

$$\log \left\{ (X_{(t+1)\delta} - X_{t\delta})^2 \right\}.$$

Nous reportons en Annexe un argument justifiant la forme de cette transformation.

En dernier lieu nous ne considérons qu'un sous-échantillon des données transformées. Formellement, en posant $\Delta = 10\delta$, pour tout $t = 1, \dots, n = 272$,

$$Y_t = \log \left\{ (X_{(t+1)\Delta} - X_{t\Delta})^2 \right\}.$$

Cette manipulation a pour principal objectif de simplifier la mise en œuvre de notre étude. En effet nous argumentons plus bas en faveur d'une méthode d'optimisation de la vraisemblance (exacte) de type quasi-Newtonienne plutôt que fondée sur l'algorithme Expectation-Maximization (EM), sa populaire alternative. La méthode quasi-Newtonienne converge beaucoup plus vite que l'algorithme EM, et les résultats qu'elle produit sont sans surprise quantitativement meilleurs (par exemple en termes de norme du gradient au terme de l'optimisation). Très aisé à programmer lorsque l'échantillon est de l'ordre de quelques centaines, le calcul de la vraisemblance exacte et de son gradient devient malheureusement très délicat lorsque celui-ci est de l'ordre de quelques milliers. Cette étude a pour ambition de poser des jalons ; nous nous emploierons dans un futur proche à adapter nos programmes pour qu'ils puissent traiter les jeux

de données originaux. Nous avons toutefois la conviction que la nature des résultats obtenus ne sera pas bouleversée.

3 Un premier modèle de Markov caché pour processus multiples

Soit Y_t^i l'observation et Z_t^i l'état caché associés au patient i , $i = 1, \dots, N$, à l'instant t , $t = 1, \dots, n$. Nous notons \mathbf{Y}^i et \mathbf{Z}^i les vecteurs n -dimensionnels des observations et états cachés du patient i . Les vecteurs nN -dimensionnels \mathbf{Y} et \mathbf{Z} contiennent l'ensemble de toutes les observations et de tous les états cachés. Les versions en lettres minuscules $y_t^i, \mathbf{y}^i, \mathbf{y}$ et $z_t^i, \mathbf{z}^i, \mathbf{z}$ correspondent à des réalisations.

3.1 Développement

Pour un nombre de régimes K fixé, les \mathbf{Z}^i sont des chaînes de Markov à valeurs dans $\{1, \dots, K\}$, de mêmes probabilités de transition $\{P_{k\ell}\}$ et loi initiale $\{\pi_k\}$. Nous imposons enfin à toutes les chaînes une initiation dans l'état numéroté 1 : pour tout i , $z_1^i = 1$, ou autrement dit $\pi_1 = 1$.

Soit $\theta \in \Theta_K$ un vecteur générique contenant tous les paramètres du modèle que nous décrivons. Sous θ , conditionnellement à \mathbf{Z} , les Y_t^i sont indépendantes, de distribution $f(y_t^i | Z_t^i = k; \theta)$ gaussienne de variance σ^2 , supposée connue, et de moyenne μ_{tk}^i ,

$$f(y_t^i | Z_t^i = k; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_t^i - \mu_{tk}^i)^2}{2\sigma^2}\right\}, \quad \text{avec}$$

$$\mu_{tk}^i = m_k + \mathbf{x}_i^T \beta_k,$$

la covariable \mathbf{x}_i valant $(0, 0)^T$ pour un patient *sain*, $(1, 0)^T$ pour un patient *vestibulaire* et $(0, 1)^T$ pour un patient *hémiplégique*. Ainsi, m_k représente la moyenne d'émission pour un patient sain dans le régime k , prise comme référence, la première et la seconde coordonnées de β_k correspondant aux modulations de cette moyenne pour les patients vestibulaires et hémiplégiques, respectivement.

Ainsi, chaque \mathbf{Y}^i est un HMM à émission gaussienne. Nous parlons de HMM pour un processus multiple car la caractérisation de la loi jointe de \mathbf{Y} requiert l'introduction des N chaînes de Markov cachées de \mathbf{Z} . Le modèle est cependant bien un HMM standard caractérisé par une matrice de transition creuse (dont la dimension dépend de N).

Finalement, la log-vraisemblance de ce modèle s'écrit

$$\ell(\theta; \mathbf{y}) = \sum_{i=1}^N \log \left\{ \sum_{\mathbf{z}^i} \pi_{z_1^i} f(y_1^i | z_1^i; \theta) \prod_{t=2}^n P_{z_{t-1}^i, z_t^i} f(y_t^i | z_t^i; \theta) \right\},$$

l'argument \mathbf{z}^i dans la somme parcourant $\{1, \dots, K\}^n$ avec les contraintes $z_1^i = 1$.

Ce modèle apparaît aussi comme un cas particulier de HMM mixte telle qu'elles sont introduites dans [2]. L'exemple 1 de la page 204 de [2] (appelé *single, patient-specific random effect*) s'écrit en effet de la même façon, à l'ajout près d'un terme d'effet aléatoire u_i centré (les u_1, \dots, u_n étant i.i.d et à densité) à la définition de μ_{ik}^i . À ce titre, les HMM mixtes sont une extension des HMM standard dans l'esprit des modèles linéaires à effets mixtes. Nous avons renoncé pour l'instant à adopter ce modèle à cause de son coût computationnel trop élevé pour notre jeu de données. À titre de comparaison, bien que les valeurs de N soient comparables dans [2] et ici, notre valeur de n est plus de dix fois supérieure à celles de [2].

Concernant le partage par toutes les chaînes de Markov cachées d'une même distribution (caractérisée par $\{P_{kl}\}$ et $\{\pi_k\}$), il relève d'une tentative de structurer de façon assez rigide le modèle que nous ajusterons. Nous pourrions invoquer aussi le rasoir d'Ockam, d'autant plus que la structure cachée du modèle (qui est une vue de l'esprit) n'est pas observée et qu'on ne peut donc pas en tester l'adéquation à une distribution (les conséquences de ce choix se lisent dans les lois marginales). À ce titre, nous pouvons argumenter sans risque que le choix d'une distribution paramétrique économe est de bonne pratique. Il faut bien voir que le choix du nombre de régimes K offre une grande latitude à la famille de modèles. Exemple parmi tant d'autres, le paramètre $\theta \in \Theta_3$ peut être choisi de telle sorte que les marginales d'émission pour les patients sains n'aient qu'une moyenne commune, que celles des patients vestibulaires en aient deux et que celles des patients hémiplésiques en aient trois. Pour conclure, le choix d'une initiation systématique des chaînes \mathbf{Z}^i dans l'état numéroté 1 fait sens. Heuristiquement, ce choix implique l'existence d'un régime (au moins) de référence correspondant à un maintien postural sans stimulation, hypothèse raisonnable du point de vue de la biomécanique.

3.2 Estimation du nombre de régimes

Le choix du nombre de régimes K est un problème crucial. La littérature dédiée à l'estimation du nombre de classes dans un mélange (dont notre problème relève) est très vaste et nous nous référons à [5, 7] pour une bibliographie. Nous proposons ici une procédure de sélection inspirée du célèbre Bayesian Information Criterion (abrégé BIC).

En nous inspirant des travaux pionniers de [11, 22] et en droite ligne du développement présenté dans [7] pour l'estimation du nombre de régimes d'un HMM à émissions gaussiennes, nous obtenons en effet un résultat de consistance que nous exposons ci-après.

Notons $(t)_+$ la partie positive du réel t . Soit

$$\begin{aligned} c_{KnN} &= \left(-K \log \frac{\Gamma(K/2)}{\Gamma(1/2)} + \frac{K^2(K-1)}{4nN} + \frac{K}{12nN} \right) \\ d_{KnN} &= \frac{3K}{2} \log \left(\frac{5}{2K} + \frac{1}{nN} \right) \end{aligned}$$

et les sommes cumulées $C_{KnN} = \sum_{k=1}^K (c_{knN})_+$, $D_{KnN} = \sum_{k=1}^K (d_{knN})_+$. Pour tout $\alpha > 2$ fixé arbitrairement, introduisons enfin le terme de pénalité

$$\text{pen}_{KnN} = \sum_{k=1}^K \left(\frac{\dim(\Theta_k) + \alpha}{2} \log nN + k \log nN \right) + C_{KnN} + D_{KnN},$$

Notons \mathbb{P}_0 la distribution de \mathbf{Y} et \mathbb{P}_θ la distribution associée au paramètre θ . Admettons que $\mathbb{P}_0 \in \cup_{K \geq 1} \{\mathbb{P}_\theta : \theta \in \Theta_K\}$ et soit K_0 l'unique entier tel que

$$\mathbb{P}_0 \in \{\mathbb{P}_\theta : \theta \in \Theta_{K_0}\} \setminus \{\mathbb{P}_\theta : \theta \in \Theta_{K_0-1}\}$$

(avec la convention $\Theta_0 = \emptyset$). On appelle K_0 l'ordre de \mathbb{P}_0 .

Proposition 1. *Sous les hypothèses énoncées dans les Sections 3.1 et 3.2, si les espaces de paramètres Θ_K sont compacts, alors \mathbb{P}_0 -presque sûrement,*

$$\widehat{K} := \min \arg \max_{K \geq 1} \left\{ \sup_{\theta \in \Theta_K} \ell(\theta; \mathbf{y}) - \text{pen}_{K n N} \right\} = K_0$$

pour n assez grand.

Cette proposition a ceci de remarquable qu'aucune borne a priori sur l'ordre de \mathbb{P}_0 n'est requise pour garantir ce résultat de consistance forte de \widehat{K} , estimateur du maximum de vraisemblance pénalisée. Comme annoncé, notre pénalisation $\text{pen}_{K n N}$ s'inspire du critère BIC, qui s'écrirait dans notre cadre sous la simple forme $\frac{1}{2} \dim(\Theta_K) \log n N$. Sous cet angle, notre pénalisation s'exprime pour partie, et principalement, comme une somme cumulée des critères BIC, et la contribution de cette somme à la pénalisation complète est un $O(K^3 \log n N)$. La seconde contribution à la définition de notre pénalisation (par ordre de grandeur relativement à K et $n N$) est la somme cumulée des $k \log n N$, dont la contribution à la pénalisation complète est un $O(K^2 \log n N)$. Il apparaîtra dans la démonstration de la Proposition 1 (reportée en Annexe) que cette seconde somme cumulée permet à la pénalisation de contrôler le maximum des carrés des observations $\max_{i \leq N, t \leq n} [Y_t^i]^2$. Il est intéressant de souligner que la nécessité de contrôler ce terme ne trouve pas à proprement parler son origine dans la technique de preuve, mais plutôt dans la nature continue des observations (exploitée dans un cadre à observations discrètes, elle ne produit pas ce terme caractéristique).

Pour conclure, notre expérience de l'utilisation du critère de type "BIC cumulé" suggère que les ordres estimés selon lui coïncident avec les ordres estimés selon le critère BIC lorsque n prend des valeurs de l'ordre de plusieurs milliers. Il a été prouvé par ailleurs dans un cadre très général de HMM [16] que le critère BIC pour l'estimation de l'ordre est consistant dès lors qu'une borne a priori sur celui-ci est connue.

4 Contribution à l'élaboration d'une notion de style postural, première partie

Préliminaire.

Pour commencer, nous estimons le paramètre σ supposé connu sur un jeu de données indépendant de celui décrit en Section 2.4. Ce jeu correspond au maintien en position debout sur une plateforme de force, les yeux fermés et en l'absence de stimulations. Nous obtenons ainsi la valeur $\sigma = 1.65$ (avec un écart-type de 0.1).

Le jeu de données décrit en Section 2.4 est découpé aléatoirement en trois parties indépendantes de tailles comparables, constituées respectivement des nombres de patients sains, vestibulaires et hémiplésiques suivant : (10, 7, 6), (10, 8, 6) et (12, 8, 4). Le premier sous-échantillon est dédié à l'estimation de l'ordre, le second à l'ajustement du modèle correspondant à l'ordre élu et le troisième étant réservé à des fins de validation.

Certes ce découpage initial rend aisé l'exposition des résultats, mais il faut tenir compte de l'aléa introduit par la répartition aléatoire en ces trois sous-échantillons. Pour ce faire nous avons ultérieurement appliqué une procédure de type bootstrap, répétant 100 fois le découpage aléatoire et les procédures intitulées ci-après *ajustement* et *validation*. Nous n'en reportons pas les résultats numériques, qui témoignent d'une bonne robustesse en termes d'ordre sélectionné (nous allons argumenter en faveur de l'estimateur \hat{K}) et d'estimateurs des paramètres d'intérêt (matrices de transition, moyennes d'émission, et probabilités de répartition telles que nous les introduirons en Section 5), les écarts-types obtenus par bootstrap étant cependant systématiquement un peu plus grands (typiquement d'une ordre de magnitude de 25%) que ceux produits par inversion de la matrice d'information.

Ajustement.

Nous reportons dans le Tableau 1 les valeurs des log-vraisemblances maximisées $\hat{\ell}_1(k)$ dans le cadre du premier modèle pour $k = 1, \dots, 10$, ainsi que leurs versions pénalisées aux sens des critères BIC et "BIC cumulé". L'ordre sélectionné par le critère BIC est $\hat{K}^{\text{BIC}} = 6$ tandis que l'ordre sélectionné par le critère "BIC cumulé" est $\hat{K} = 5$.

Lequel des deux choisir ? Nous ajustons les modèles d'ordre 5 et 6 au second sous-échantillon. L'inspection des résultats apporte une réponse catégorique à la question du choix. L'estimateur du maximum de vraisemblance dans $\Theta_{\hat{K}^{\text{BIC}}}$ présente deux valeurs aberrantes (correspondant à $\{\beta_k\}$) et la matrice d'information observée n'est pas inversible. En revanche, l'ajustement dans $\Theta_{\hat{K}}$ fait sens, et la matrice d'information observée est inversible. Nous adoptons donc $\hat{K} = 5$ et exposons ses résultats dans le Tableau 3. La Figure 1 présente visuellement les différentes moyennes d'émission.

On notera que le premier régime des patients sains diffère significativement des premiers régimes des autres patients, par ailleurs non significativement différents. Dans ce régime, le comportement des patients sains apparaît beaucoup plus concentré (au sens de la concentration d'une variable aléatoire autour de sa moyenne) que ceux des autres patients. En d'autres termes, les comportements des patients vestibulaires et hémiplésiques apparaissent beaucoup plus dispersés que celui des patients sains.

On notera aussi que les comportements des patients sains et des patients hémiplésiques dans le régime 3 diffèrent significativement, et que les comportements des patients sains et vestibulaires diffèrent significativement dans le régime 4.

Curiosité enfin, les moyennes d'émission des trois types de patients ne sont pas significativement différentes dans les régimes 2 et 5, et ces moyennes ne varient pas significativement d'un régime à l'autre. On lit dans la matrice de transition estimée que le régime 2 est de nature transitoire vers le régime 5 (probabilité $\sim 90\%$) et le régime 3 (probabilité $\sim 10\%$).

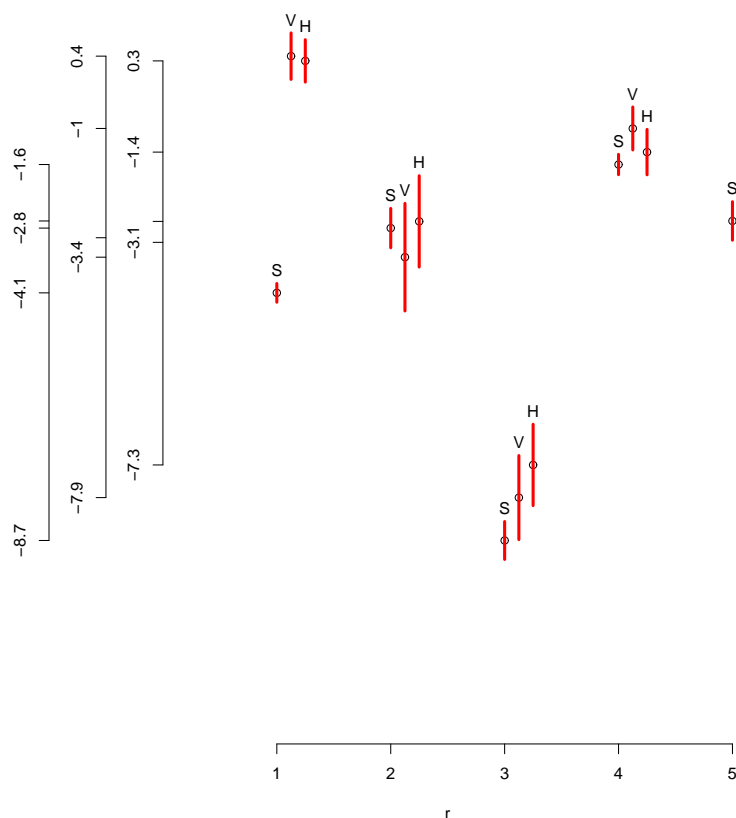


FIGURE 1 – Les différentes moyennes d'émission du premier modèle d'ordre $\hat{K} = 5$ ajusté. Les segments verticaux représentent les intervalles de confiance de niveau 95% des moyennes d'émission pour chaque régime et chaque type de patient. Les segments correspondant aux patients sains, vestibulaires et hémiplegiques sont respectivement surmontés d'un S, d'un V et d'un H. L'estimée ponctuelle des moyennes est indiquée par un cercle. Les barres verticales situées à gauche de la figure permettent d'évaluer l'échelle numérique. Les barres correspondent, de gauche à droite, aux estimées ponctuelles des patients sains, vestibulaires et hémiplegiques respectivement.

Validation.

Pour juger de la qualité de l'ajustement, nous proposons d'essayer de classer chaque membre du sous-échantillon de validation dans l'un des trois groupes de patients à partir des seules trajectoires observées.

Sur les 12 patients sains, 8 sont identifiés comme tels, 2 comme vestibulaires et 2 comme hémiplegiques (66% d'identification correcte). Sur les 8 patients vestibulaires, 6 sont identifiés comme tels, 1 comme sain et 1 comme hémiplegique (75% d'identification correcte). Sur les 4 patients hémiplegiques, deux sont identifiés comme tels, 1 comme sain et 1 comme vestibulaire (50% d'identification correcte). Dans l'ensemble, 66% des patients sont correctement identifiés.

Nous jugeons décevant ce résultat. Peut-être cette performance en demi-teinte est-elle un indice de ce que les types de patients ne caractérisent pas la nature des trajectoires observées. Nous proposons dans la prochaine section un second modèle ayant vocation à explorer la voie que cette remarque ouvre.

5 Un second modèle de Markov caché pour processus multiples

Nous avons argumenté dans la Section 4 que le premier modèle de Markov caché pour processus multiples construit en Section 3 laisse à penser que les trois groupes de patients (sains, codé par $\mathbf{x}^T = (0, 0)$; vestibulaires, codés par $\mathbf{x}^T = (1, 0)$; hémiplegiques, codé par $\mathbf{x}^T = (0, 1)$) ne sont pas associés chacun à un comportement caractéristique (en termes de réponse au protocole médical décrit dans la Section 2.3). Nous proposons dans cette section une extension du premier modèle fondée sur l'idée *qu'il existe bien des comportements différents mais que ceux-ci ne sont pas rigoureusement caractéristiques des groupes de patients*.

Plus formellement, nous supposons que chaque patient i , $i = 1, \dots, N$, a associée une variable aléatoire cachée W^i à valeurs dans $\{1, 2, 3\}$. Selon ce nouveau modèle, il existe jusqu'à trois types de comportements et $W^i = w$ si le patient i est un représentant du groupe w .

Le nouveau modèle est présenté en détails dans la prochaine section. La question de l'estimation de l'ordre est l'objet de la section suivante. Nous notons dans la suite de cet article \mathbf{W} le vecteur N -dimensionnel des variables cachées W^i et \mathbf{w} , w^i des réalisations respectives.

5.1 Développement

Soit $\psi = (\psi_{\mathbf{v}w} : \mathbf{v} \in \{(0, 0)^T, (1, 0)^T, (0, 1)^T\}, w = 1, 2, 3) \in \Psi$ un paramètre additionnel à coordonnées positives satisfaisant l'égalité $\sum_{w=1}^3 \psi_{\mathbf{v}w} = 1$ pour tout \mathbf{v} . Le paramètre ψ caractérise la distribution de \mathbf{W} selon les groupes auxquels les patients appartiennent (c'est-à-dire selon les \mathbf{x}_i). Nous supposons que les W^i sont mutuellement indépendantes et, sous $\psi \in \Psi$ un paramètre générique, que la probabilité pour que le patient i appartienne au groupe w vaut $\psi_{\mathbf{x}_i w}$.

Formellement,

$$\mathbb{P}_\psi(\mathbf{W} = \mathbf{w}) = \prod_{i=1}^N \psi_{\mathbf{x}_i w^i}.$$

Désormais, sous $(\theta, \psi) \in \Theta_K \times \Psi$ un paramètre générique d'un modèle d'ordre au plus K , conditionnellement à $\mathbf{W} = \mathbf{w}$ et à \mathbf{Z} , les Y_t^i sont indépendantes, de distribution $f(y_t^i | Z_t^i = k, w^i; \theta)$ gaussienne de variance σ^2 et de moyenne

$$\mu_{tk}^i = \begin{cases} m_k & \text{si } w^i = 1 \\ m_k + (1, 0)^T \beta_k & \text{si } w^i = 2 \\ m_k + (0, 1)^T \beta_k & \text{si } w^i = 3 \end{cases}.$$

La log-vraisemblance de ce second modèle a donc la forme

$$\ell(\theta, \psi; \mathbf{y}) = \sum_{i=1}^N \log \left\{ \sum_{w^i} \psi_{\mathbf{x}_i w^i} \left(\sum_{z^i} \pi_{z^i} f(y_1^i | z_1^i, w^i; \theta) \prod_{t=2}^n P_{z_{t-1}^i, z_t^i} f(y_t^i | z_t^i, w^i; \theta) \right) \right\},$$

les arguments w^i et z^i dans les sommes parcourant respectivement $\{1, 2, 3\}$ et $\{1, \dots, K\}^n$ avec les contraintes $z_1^i = 1$.

Le modèle apparaît comme un mélange de modèles similaires à celui introduit en Section 3. Bien évidemment, $\ell(\theta, \psi^0; \mathbf{y}) = \ell(\theta; \mathbf{y})$ pour tout $\theta \in \Theta_K$ si ψ^0 satisfait $\psi_{\mathbf{v}_1}^0 = 1$ pour $\mathbf{v} = (0, 0)^T$, $\psi_{\mathbf{v}_2}^0 = 1$ pour $\mathbf{v} = (1, 0)^T$, $\psi_{\mathbf{v}_3}^0 = 1$ pour $\mathbf{v} = (0, 1)^T$ — c'est-à-dire si les comportements sont rigoureusement caractérisés par les groupes d'appartenance. On peut noter que le modèle ainsi construit s'écrit encore comme un cas particulier de l'exemple 1 de la page 204 de [2].

5.2 Estimation du nombre de régimes

On note toujours \mathbb{P}_0 la distribution de \mathbf{Y} et $\mathbb{P}_{\theta, \psi}$ la distribution associée au paramètre (θ, ψ) . Admettons que $\mathbb{P}_0 \in \cup_{K \geq 1} \{\mathbb{P}_{\theta, \psi} : \theta \in \Theta_K, \psi \in \Psi\}$ et soit K_0 l'unique entier tel que

$$\mathbb{P}_0 \in \{\mathbb{P}_{\theta, \psi} : \theta \in \Theta_{K_0}, \psi \in \Psi\} \setminus \{\mathbb{P}_{\theta, \psi} : \theta \in \Theta_{K_0-1}, \psi \in \Psi\}$$

(avec la convention $\Theta_0 = \emptyset$).

Il apparaît qu'une modification mineure de la pénalité introduite en Section 7.3 permet d'obtenir un résultat similaire à la Proposition 1. Sa démonstration est reportée en Annexe.

Soit donc le terme de pénalité complémentaire

$$\text{comp}_{K_n N} = K \left(\frac{\dim(\Psi)}{2} \log N + 3 \log 2 \right).$$

Proposition 2. *Sous les hypothèses énoncées dans les Sections 5.1 et 5.2, si les espaces de paramètres Θ_K sont compacts, alors \mathbb{P}_0 -presque sûrement,*

$$\widehat{K} := \min_{K \geq 1} \arg \max \left\{ \sup_{\theta \in \Theta_K} \sup_{\psi \in \Psi} \ell(\theta, \psi; \mathbf{y}) - \text{pen}_{K_n N} - \text{comp}_{K_n N} \right\} = K_0$$

pour n assez grand.

Les commentaires qui étaient pertinents pour le terme de pénalité $\text{pen}_{K n N}$ dans le cadre du premier modèle le sont encore pour le terme de pénalité complété (c'est-à-dire ajusté au cadre du second modèle) $\text{pen}_{K n N} + \text{comp}_{K n N}$. Il est bon de noter que le critère BIC s'écrivait ici sous la simple forme $\frac{1}{2} \dim(\Theta_K) \log n N + \frac{1}{2} \dim(\Psi) \log N$ et qu'à ce titre, la pénalité complétée apparaissant dans la Proposition 2 s'exprime elle aussi pour partie, et principalement, comme une somme cumulée des critères BIC.

6 Contribution à l'élaboration d'une notion de style postural, seconde partie

Ajustement.

Nous reportons dans le Tableau 2 les valeurs des log-vraisemblances maximisées $\hat{\ell}_2(k)$ dans le cadre du second modèle pour $k = 1, \dots, 10$, ainsi que leurs versions pénalisées aux sens des critères BIC et "BIC cumulé". L'ordre sélectionné par le critère BIC est $\hat{K}^{\text{BIC}} = 6$ tandis que l'ordre sélectionné par le critère "BIC cumulé" est $\hat{K} = 3$.

La question du choix se pose de nouveau. Nous y répondons dans les mêmes termes. Nous ajustons les modèles d'ordre 3 et 6 au second sous-échantillon. L'inspection des résultats apporte une réponse catégorique à la question du choix. L'estimateur du maximum de vraisemblance dans $\Theta_{\hat{K}^{\text{BIC}}}$ présente plusieurs valeurs aberrantes et la matrice d'information observée n'est pas inversible. En revanche, l'ajustement dans $\Theta_{\hat{K}}$ fait sens, et la matrice d'information observée est inversible. Nous adoptons donc cette fois encore $\hat{K} = 3$ et exposons ses résultats dans le Tableau 4. La Figure 2 présente visuellement les différentes moyennes d'émission.

On notera que pour chaque régime, les moyennes d'émissions sont classées dans le même ordre : pour tous les régimes, le premier comportement est celui de la moindre dispersion, le second comportement est intermédiaire et le troisième comportement est caractérisé par la plus grande dispersion des émissions. Dans le premier régime, la moyenne du premier comportement est significativement différente des deux autres moyennes. Dans le second régime, les trois moyennes sont significativement différentes. Dans le troisième régime, la moyenne du premier comportement est significativement différente de celle du troisième.

Enfin, la lecture de l'estimée des probabilités $\{\hat{\psi}_{vw}\}$ de répartition dans chaque groupe de comportement selon la nature du patient permet d'identifier le premier des comportements comme celui des patients sains, le second comme celui des patients vestibulaires plutôt que hémiplegiques (75% contre 25%) et le troisième comme celui des patients hémiplegiques plutôt que vestibulaires (66% contre 33%).

Prolongements.

Ce travail nous ouvre de nombreuses perspectives. Nous souhaitons bien sûr mettre au point la procédure qui permettra d'attribuer un comportement parmi les trois à un individu sur la base de la trajectoire observée seule. L'objectif à un peu plus long terme est de pouvoir procéder à un regroupement selon les trois comportements d'un ensemble de patients. Par ailleurs, on

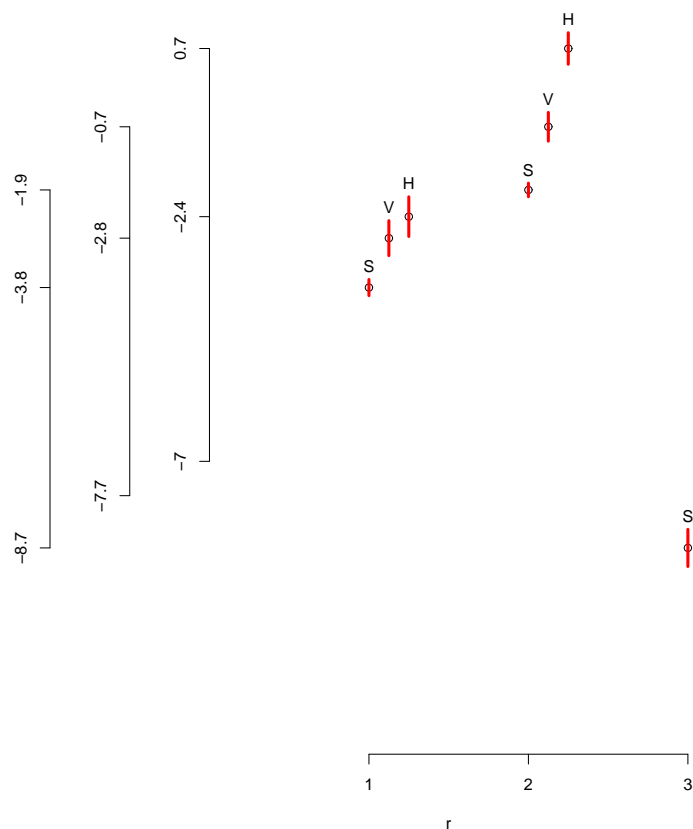


FIGURE 2 – Les différentes moyennes d'émission du second modèle d'ordre $\hat{K} = 3$ ajusté. Les segments verticaux représentent les intervalles de confiance de niveau 95% des moyennes d'émission pour chaque régime et chaque type de comportement. Les segments correspondant aux premier, second et troisième comportements sont respectivement surmontés d'un a, d'un b et d'un c. L'estimée ponctuelle des moyennes est indiquée par un cercle. Les barres verticales situées à gauche de la figure permettent d'évaluer l'échelle numérique. Les barres correspondent, de gauche à droite, aux estimées ponctuelles des comportements a, b, c respectivement.

serait en droit de nous reprocher de ne pas utiliser les autres covariables dont nous disposons : âge, sexe, taille, poids, profession. Ce sera aussi l'une de nos priorités dans un futur proche. Nous souhaitons enfin bien sûr appliquer cette approche à l'ensemble des trajectoires dont nous disposons et telles que nous les observons (c'est-à-dire sans la substitution de Δ à δ).

7 Annexe

L'Annexe est décomposée en trois sections : dans la première, nous exposons un argument qui justifie l'intérêt que nous portons à $Y_t = \log\{(X_{t+1} - X_t)^2\}$; la seconde est consacrée à la démonstration de la Proposition 1 et la troisième à celle de la Proposition 2.

7.1 Commentaire de la définition des Y_t^i

Dans cet article, nous accordons notre attention à $\{Y_t\} = \{\log[(X_{t+1} - X_t)^2]\}$, que nous considérons comme un représentant de la volatilité du processus $\{X_t\}$. Nous développons ci-dessous, dans un modèle non paramétrique de diffusion, un argument justifiant cette interprétation. Bien que très général, ce cadre n'est pas compatible avec les modélisations que nous avons construites dans l'article. Nous pensons néanmoins que l'argument éclaire notre démarche.

Un argument éclairant.

Supposons donc (exclusivement dans cette section) que le processus $\{X_t\}$ satisfait l'équation différentielle stochastique

$$dX_t = X_0 + b(X_t)dt + \sigma(X_t)dB_t,$$

où $\{B_t\}$ est un mouvement brownien, X_0 est une variable aléatoire indépendante de $\{B_t\}$, les fonctions de drift b et de volatilité σ étant de classe \mathcal{C}^2 et lipschitziennes sur \mathbb{R}_+ . En particulier, il existe une constante $K > 0$ telle que, pour tout x, y , $|\sigma(x) - \sigma(y)| \leq K|x - y|$ et $b^2(x) + \sigma^2(x) \leq K(1 + x^2)$. La théorie des diffusions garantit l'existence d'une unique solution adaptée à $\{\sigma(X_0, B_s : s \leq t)\}$. Pour simplifier, nous supposons enfin que le processus $\{X_t\}$ est stationnaire et qu'il existe $\sigma_0 > 0$ telle que $\sigma(x) \geq \sigma_0$ pour tout x .

Soit Δ un pas d'échantillonnage (petit). Alors

$$\begin{aligned}
\left(\frac{X_{(i+1)\Delta} - X_{i\Delta}}{\sqrt{\Delta}}\right)^2 &= \Delta^{-1} \left(\int_{i\Delta}^{(i+1)\Delta} b(X_s) ds + \int_{i\Delta}^{(i+1)\Delta} \sigma(X_s) dB_s \right)^2 \\
&= \Delta^{-1} \left\{ \left(\int_{i\Delta}^{(i+1)\Delta} b(X_s) ds \right)^2 + \left(\int_{i\Delta}^{(i+1)\Delta} \sigma(X_s) dB_s \right)^2 \right. \\
&\quad \left. + 2 \left(\int_{i\Delta}^{(i+1)\Delta} b(X_s) ds \right) \left(\int_{i\Delta}^{(i+1)\Delta} \sigma(X_s) dB_s \right) \right\} \\
&= \sigma^2(X_{i\Delta}) \left(\frac{B_{(i+1)\Delta} - B_{i\Delta}}{\sqrt{\Delta}} \right)^2 + R_i, \tag{1}
\end{aligned}$$

pour un certain R_i qui satisfait, nous le montrons ci-dessous, $E(|R_i|) = O(\sqrt{\Delta})$. Posons $\varepsilon_i = \Delta^{-1/2}(B_{(i+1)\Delta} - B_{i\Delta})$, variable aléatoire de loi normale centrée réduite. Alors

$$\log \left\{ \left(\frac{X_{(i+1)\Delta} - X_{i\Delta}}{\sqrt{\Delta}} \right)^2 \right\} = \log\{\sigma^2(X_{i\Delta})\} + \log \left\{ \varepsilon_i^2 + \frac{R_i}{\sigma(X_{i\Delta})} \right\},$$

le second terme convergeant en probabilité vers $\log\{\varepsilon_i^2\}$ quand Δ tend vers 0. Ainsi, la formule précédente met en évidence la relation entre le processus $\{Y_t\}$ et la volatilité de $\{X_t\}$ dans notre étude.

Démonstration de l'égalité (1).

Pour commencer, le terme dominant provient de la décomposition suivante :

$$\begin{aligned}
\Delta^{-1} \left(\int_{i\Delta}^{(i+1)\Delta} \sigma(X_s) dB_s \right)^2 &= \sigma^2(X_{i\Delta}) \varepsilon_i^2 + \Delta^{-1} \left(\int_{i\Delta}^{(i+1)\Delta} [\sigma(X_s) - \sigma(X_{i\Delta})] dB_s \right)^2 \\
&\quad + 2\Delta^{-1/2} \varepsilon_i \left(\int_{i\Delta}^{(i+1)\Delta} [\sigma(X_s) - \sigma(X_{i\Delta})] dB_s \right) = \sigma^2(X_{i\Delta}) \varepsilon_i^2 + R_{i,1} + R_{i,2}
\end{aligned}$$

avec $R_{i,1} \geq 0$. En effet

$$\begin{aligned}
E(|R_{i,1}|) &= \Delta^{-1} \int_{i\Delta}^{(i+1)\Delta} E([\sigma(X_s) - \sigma(X_{i\Delta})]^2) ds \\
&\leq K^2 \Delta^{-1} \int_{i\Delta}^{(i+1)\Delta} E([X_s - X_{i\Delta}]^2) ds. \tag{2}
\end{aligned}$$

Or pour tout $s \in [i\Delta, (i+1)\Delta]$, grâce à l'inégalité $\frac{1}{2}(a+b)^2 \leq (a^2 + b^2)$,

$$\frac{1}{2} E([X_s - X_{i\Delta}]^2) \leq E \left[\left(\int_{i\Delta}^s b(X_u) du \right)^2 \right] + E \left[\left(\int_{i\Delta}^s \sigma(X_u) dB_u \right)^2 \right], \tag{3}$$

où, en vertu de l'inégalité de Cauchy-Schwarz et de $b^2(x) \leq K(1+x^2)$,

$$E \left[\left(\int_{i\Delta}^s b(X_u) du \right)^2 \right] \leq (s - i\Delta) E \left(\int_{i\Delta}^s b^2(X_u) du \right) \leq K\Delta^2(1 + E(X_0^2)), \quad (4)$$

tandis que

$$E \left[\left(\int_{i\Delta}^s \sigma(X_u) dB_u \right)^2 \right] = E \left(\int_{i\Delta}^s \sigma^2(X_u) du \right) \leq K\Delta(1 + E(X_0^2)). \quad (5)$$

En injectant (4) et (5) dans (3) puis l'inégalité obtenue dans (2), on obtient que $E(|R_{i,1}|) = O(\Delta)$. Par ailleurs, l'inégalité de Cauchy-Schwarz donne

$$[E(|R_{i,2}|)]^2 \leq 4E(\varepsilon_i^2)E(|R_{i,1}|),$$

dont on déduit évidemment que $E(|R_{i,2}|) = O(\sqrt{\Delta})$.

Le terme résiduel positif $R_{i,3} = \Delta^{-1} \left(\int_{i\Delta}^{(i+1)\Delta} b^2(X_s) ds \right)^2$ est à son tour contrôlé en espérance grâce à (4), avec $E(|R_{i,3}|) = O(\Delta)$. Quant au dernier terme résiduel

$$R_{i,4} = 2\Delta^{-1} \left(\int_{i\Delta}^{(i+1)\Delta} b(X_s) ds \right) \left(\int_{i\Delta}^{(i+1)\Delta} \sigma(X_s) dB_s \right),$$

nous invoquons une encore l'inégalité de Cauchy-Schwarz pour obtenir

$$[E(|R_{i,4}|)]^2 \leq 4\Delta^{-2} E \left[\left(\int_{i\Delta}^s b(X_u) du \right)^2 \right] E \left[\left(\int_{i\Delta}^s \sigma(X_u) dB_u \right)^2 \right],$$

les inégalités (4) et (5) garantissant alors que $E(|R_{i,4}|) = O(\sqrt{\Delta})$.

En résumé, le terme résiduel $R_i = R_{i,1} + R_{i,2} + R_{i,3} + R_{i,4}$ satisfait bien $E(|R_i|) = O(\sqrt{\Delta})$, comme annoncé.

7.2 Aspects numériques

Le plus souvent, les paramètres d'un HMM sont estimés en utilisant l'algorithme Expectation-Maximization (EM). Nous lui préférons ici un algorithme de maximisation de la vraisemblance *exacte*, calculée à partir de son expression faisant intervenir des produits de matrices.

Soit A^{i1} le vecteur de coordonnées $\pi_k f(y_1^i | k; \theta)$ et A^{it} la matrice d'entrées $P_{k\ell} f(y_t^i | k; \theta)$ pour $t = 2, \dots, n$. On note $\mathbf{1}$ le vecteur K -dimensionnel à coordonnées toutes égales à 1. La log-vraisemblance du premier modèle s'écrit de façon équivalente sous la forme

$$\ell(\theta; \mathbf{y}) = \sum_{i=1}^n \log \left\{ (A^{i1})^T \prod_{t=2}^n A^{it} \mathbf{1} \right\}.$$

Les matrices ne commutent pas, il faut bien sûr lire le produit comme $A^{i2} \times \dots \times A^{in}$.

On peut aussi calculer de façon exacte le score.

Soit en effet $B^{i1} = \mathbf{I}$, la matrice identité $K \times K$, et B^{is} le produit de matrices $\prod_{t=2}^s A^{it}$ pour $s = 2, \dots, n$. Soit ϑ l'une quelconque des coordonnées de θ et ∂_ϑ l'opérateur de dérivation par rapport à ϑ . La relation de récurrence suivante, valable pour tout $s \geq 2$,

$$\partial_\vartheta\{B^{is}\} = \partial_\vartheta\{B^{i(s-1)}A^{is}\} = \partial_\vartheta\{B^{i(s-1)}\}A^{is} + B^{i(s-1)}\partial_\vartheta\{A^{is}\}$$

permet d'obtenir la dérivée $\partial_\vartheta\{B^{in}\}$ à partir des expressions, simples, des $\partial_\vartheta\{A^{it}\}$. Il vient alors immédiatement que

$$\partial_\vartheta\ell(\theta; \mathbf{y}) = \sum_{i=1}^n \frac{(\partial_\vartheta\{A^{i1}\})^T B^{in} + (A^{i1})^T \partial_\vartheta\{B^{in}\}}{(A^{i1})^T B^{in} \mathbf{1}}.$$

Cette approche numérique s'adapte sans aucune difficulté à la forme de la log-vraisemblance du second modèle.

La maximisation des vraisemblances est donc menée, à partir de plusieurs valeurs initiales, en appliquant la méthode quasi-Newtonienne BFGS [24] telle qu'elle est implémentée dans R [27].

7.3 Preuve de la Proposition 1

La Proposition 1 est un corollaire de l'inégalité dite "de mélange" exposée dans la Proposition 3, qui consiste en une comparaison entre la log-vraisemblance maximisée et la statistique de mélange définie ci-après.

Par souci de simplicité, nous utiliserons par la suite la notation générique $f(u)$ pour la densité (ou la masse discrète) de la variable aléatoire (ou du vecteur) U , et $f(u|v)$ pour la version conditionnelle à $V = v$. À titre d'exemple, on a ainsi $f(z_{t+1}^i | z_t^i; \theta) = P_{z_t^i z_{t+1}^i}$.

Soit $\tau > 0$ une constante à choisir ultérieurement et K un nombre de régimes fixé. Quitte à procéder à un changement de paramétrisation, on peut supposer ici que le paramètre θ consiste en les $\{P_{k\ell}\}$, $\{\pi_k\}$, $\{m_k^1\}$ (les K moyennes d'émission pour les patients sains, associés à $w = 1$), $\{m_k^2\}$ (celles pour les patients vestibulaires, associés à $w = 2$) et $\{m_k^3\}$ (celles pour les patients hémiplegiques, associés à $w = 3$). Soit ν_K la loi a priori sur Θ_K caractérisée par les propriétés suivantes :

- sous ν_K , $\{P_{k\ell}\}$, $\{m_k^1\}$, $\{m_k^2\}$ et $\{m_k^3\}$ sont indépendantes ;
- sous ν_K , $\pi_1 = 1$ (les chaînes de Markov \mathbf{Z}^i sont toutes initiées en 1) ;
- sous ν_K , les vecteurs $\{P_{1\ell}\}, \dots, \{P_{K\ell}\}$ sont i.i.d de loi Dirichlet($1/2, \dots, 1/2$) ;
- sous ν_K , m_1^w, \dots, m_K^w sont i.i.d de loi normale centrée de variance τ^2 pour $w = 1, 2, 3$.

La statistique de mélange d'ordre K est enfin définie par

$$q_K(\mathbf{Y}) = \int_{\Theta_K} f(\mathbf{Y}; \theta) \nu_K(d\theta). \quad (6)$$

Proposition 3. Pour tout nombre de régimes $K \geq 1$,

$$0 \leq \sup_{\theta \in \Theta_K} \log \frac{f(\mathbf{Y}; \theta)}{q_K(\mathbf{Y})} \leq \frac{\dim(\Theta_K)}{2} \log nN + \frac{3K}{2\tau^2} \max_{i \leq N, t \leq n} [Y_t^i]^2 + c_{KnN} + \frac{3K}{2} \log \left(\frac{\tau^2}{3K\sigma^2} + \frac{1}{nN} \right). \quad (7)$$

Introduisons quelques notations pour l'exposition de la démonstration de la Proposition 3. Les statistiques $q_K(\mathbf{y}|\mathbf{z})$ et $q_K(\mathbf{z})$ sont définies à partir de (6) en substituant à $f(\mathbf{Y}; \theta)$ soit $f(\mathbf{y}|\mathbf{z}; \theta)$, soit $f(\mathbf{z}; \theta)$, respectivement. On note I^w l'ensemble des indices i tels que le patient i soit de type w , $T_k^i = \{t \leq n : z_t^i = k\}$, $N_k^w = \sum_{i \in I^w} \text{card}(T_k^i)$ et \bar{y}_k^w la moyenne empirique des y_t^i pour $i \in I^w$ et $t \in T_k^i$.

Démonstration de la Proposition 3. Grâce à l'inégalité $\sum_{k \leq K} \alpha_k / \sum_{k \leq K} \beta_k \leq \max_{k \leq K} \alpha_k / \beta_k$ (valable pour tout $\{\alpha_k\}$ positifs et tout $\{\beta_k\}$ strictement positifs), et puisque

$$f(\mathbf{y}; \theta) = \sum_{\mathbf{z}} f(\mathbf{y}|\mathbf{z}; \theta) f(\mathbf{z}; \theta) \quad \text{et} \quad q_K(\mathbf{y}) = \sum_{\mathbf{z}} q_K(\mathbf{y}|\mathbf{z}) q_K(\mathbf{z})$$

(l'argument \mathbf{z} dans les deux sommes précédentes et dans la suite de la démonstration parcourt $(\{1, \dots, K\}^n)^N$ avec les conditions initiales $z_1^i = 1$), il apparaît que

$$\sup_{\theta \in \Theta_K} \log \frac{f(\mathbf{y}; \theta)}{q_K(\mathbf{y})} \leq \sup_{\theta \in \Theta_K} \max_{\mathbf{z}} \left\{ \log \frac{f(\mathbf{z}; \theta)}{q_K(\mathbf{z})} + \log \frac{f(\mathbf{y}|\mathbf{z}; \theta)}{q_K(\mathbf{y}|\mathbf{z})} \right\}. \quad (8)$$

Chacun des deux termes requiert un traitement spécifique. On obtient d'une part que

$$\sup_{\theta \in \Theta_K} \max_{\mathbf{z}} \log \frac{f(\mathbf{z}; \theta)}{q_K(\mathbf{z})} \leq \frac{K(K-1)}{2} \log nN - K \log \frac{\Gamma(K/2)}{\Gamma(1/2)} + \frac{K^2(K-1)}{4nN} + \frac{K}{12nN}, \quad (9)$$

grâce au choix de l'a priori sur $\{P_{k\ell}\}$ (voir [10]) et à une approximation de Robbins-Stirling. D'autre part, à \mathbf{z} fixé,

$$f(\mathbf{y}|\mathbf{z}; \theta) \leq f(\mathbf{y}|\mathbf{z}; \hat{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}^{nN}} \prod_{w=1}^3 \prod_{k=1}^K \exp \left\{ -\frac{1}{2\sigma^2} \left(\sum_{i \in I^w, t \in T_k^i} [y_t^i]^2 - N_k^w [\bar{y}_k^w]^2 \right) \right\} \quad (10)$$

uniformément en $\theta \in \Theta_K$ ($\hat{\theta}$ est l'estimateur du maximum de vraisemblance), tandis qu'un calcul sans difficulté particulière conduit à l'égalité

$$q_K(\mathbf{y}|\mathbf{z}) = \frac{1}{\sqrt{2\pi\sigma^2}^{nN}} \prod_{w=1}^3 \prod_{k=1}^K \frac{1}{\sqrt{1 + \frac{\tau^2 N_k^w}{\sigma^2}}} \times \exp \left\{ -\frac{1}{2\sigma^2} \left(\sum_{i \in I^w, t \in T_k^i} [y_t^i]^2 - \frac{N_k^w [\bar{y}_k^w]^2}{1 + \frac{\sigma^2}{\tau^2 N_k^w}} \right) \right\}. \quad (11)$$

En combinant (10) et (11), il vient que

$$\begin{aligned}
\log \frac{f(\mathbf{y}|\mathbf{z}; \theta)}{q_K(\mathbf{y}|\mathbf{z})} &\leq \sum_{w=1}^3 \sum_{k=1}^K \left\{ \frac{1}{2} \log \left(1 + \frac{\tau^2 N_k^w}{\sigma^2} \right) + \frac{1}{2\sigma^2} \frac{N_k^w [\bar{y}_k^w]^2}{1 + \frac{\tau^2 N_k^w}{\sigma^2}} \right\} \\
&\leq \sum_{w=1}^3 \sum_{k=1}^K \frac{1}{2} \log \left(1 + \frac{\tau^2 N_k^w}{\sigma^2} \right) + \frac{3K}{2\tau^2} \max_{i \leq N, t \leq n} [y_t^i]^2 \\
&\leq \frac{3K}{2} \log \left(1 + \frac{\tau^2 nN}{\sigma^2 3K} \right) + \frac{3K}{2\tau^2} \max_{i \leq N, t \leq n} [y_t^i]^2, \tag{12}
\end{aligned}$$

la dernière inégalité découlant d'un argument de convexité.

Le résultat final est le fruit de (9) et (12). \square

Il est important de noter que ce sont les deux premiers termes du membre de droite de l'inégalité (7) qui dominent le majorant. Ils le dominent d'ailleurs avec le même rythme par rapport à nN , puisque (voir le Lemme 3 de [7]), dans notre cadre de travail :

$$\mathbb{P}_0 \left\{ \max_{i \leq N, t \leq n} [Y_t^i]^2 \geq 5\sigma^2 \log nN \right\} \leq (nN)^{-3/2}. \tag{13}$$

Ainsi, le membre de droite de (13) étant le terme général d'une série en n convergente, le lemme de Borel-Cantelli garantit que \mathbb{P}_0 -presque sûrement, $\max_{i \leq N, t \leq n} [Y_t^i]^2 \leq 5\sigma^2 \log nN$ pour n assez grand.

Nous sommes désormais en mesure de démontrer que

Proposition 4. \mathbb{P}_0 -presque sûrement, $\widehat{K} \leq K_0$ pour n assez grand.

Démonstration. Soit Ω_{nN} l'événement dont la probabilité est majorée en (13). Nous avons déjà argumenté que $\mathbb{P}_0\{\limsup_n \Omega_{nN}^c\} = 0$, donc il nous suffit de considérer la probabilité de l'événement $[\widehat{K} > K_0] \cap \Omega_{nN}^c$.

Soit $K > K_0$ et $\mathbb{P}_0 = \mathbb{P}_{\theta_0}$ pour $\theta_0 \in \Theta_{K_0}$. Si $\widehat{K} = K$, alors

$$\begin{aligned}
\ell(\theta_0; \mathbf{Y}) &\leq \sup_{\theta \in \Theta_{K_0}} \ell(\theta; \mathbf{Y}) \leq \sup_{\theta \in \Theta_K} \ell(\theta; \mathbf{Y}) + \text{pen}_{K_0 nN} - \text{pen}_{K nN} \\
&\leq \log q_K(\mathbf{Y}) + \Delta_{K nN}
\end{aligned}$$

en vertu de la Proposition 3 (avec $\tau^2 = \frac{15\sigma^2}{2}$ — notez que cette valeur n'a aucune signification particulière), où

$$\begin{aligned}
\Delta_{K nN} &= \text{pen}_{K_0 nN} - \text{pen}_{K nN} + \frac{\dim(\Theta_K)}{2} \log nN + \frac{K}{5\sigma^2} \max_{i \leq N, t \leq n} [Y_t^i]^2 + c_{K nN} + d_{K nN} \\
&\leq -\frac{\alpha(K - K_0)}{2} \log nN + \frac{K}{5\sigma^2} \left(\max_{i \leq N, t \leq n} [Y_t^i]^2 - 5\sigma^2 \log nN \right).
\end{aligned}$$

Ainsi, sur Ω_{nN}^c , $\Delta_{KnN} \leq -\frac{\alpha(K-K_0)}{2} \log nN$. L'argument principal est un changement de probabilité :

$$\begin{aligned} \mathbb{P}_0\{\widehat{K} = K \cap \Omega_{nN}\} &= \int \frac{f(\mathbf{y}; \theta_0)}{q_K(\mathbf{y})} \mathbb{1}\{\widehat{K} = K \cap \Omega_{nN}\} q_K(\mathbf{y}) d\mathbf{y} \\ &\leq \int \exp\left\{-\frac{\alpha(K-K_0)}{2} \log nN\right\} q_K(\mathbf{y}) d\mathbf{y} = (nN)^{-\alpha(K-K_0)/2}. \end{aligned}$$

En conséquence,

$$\mathbb{P}_0\{\widehat{K} > K_0 \cap \Omega_{nN}\} = O((nN)^{-\alpha/2}),$$

et, α étant opportunément choisi supérieur à 2, une seconde application du lemme de Borel-Cantelli permet de conclure. \square

En complément, il s'avère aussi qu'en vertu d'un résultat de type loi des grands nombres,

Proposition 5. \mathbb{P}_0 -presque sûrement, $\widehat{K} \geq K_0$ pour n assez grand.

Démonstration. Il suffit bien entendu de prouver que pour tout $K < K_0$, $\mathbb{P}_0\{\widehat{K} = K\} = 0$ pour n assez grand. Or, si $\widehat{K} = K < K_0$, alors $\sup_{\theta \in \Theta_K} \ell(\theta; \mathbf{Y}) \geq \ell(\theta_0; \mathbf{Y}) + o(n)$ (avec $\theta_0 \in \Theta_{K_0}$ tel que $\mathbb{P}_{\theta_0} = \mathbb{P}_0$). Par ailleurs, $\theta \mapsto \ell(\theta; \mathbf{y})$ est uniformément continue sur Θ_K , le module de continuité pouvant de pas dépendre de \mathbf{y} . L'espace Θ_K étant compact, il existe pour tout $\varepsilon > 0$ une collection finie et déterministe $\{\theta_j\} \subset \Theta_K$ telle que $\sup_{\theta \in \Theta_K} \ell(\theta; \mathbf{Y}) \leq \max_j \ell(\theta_j; \mathbf{Y}) + n\varepsilon$.

Maintenant, le Lemme 8 de [7] (un résultat à la Shannon-Breiman-McMillan, lui-même dérivé des Théorèmes 2 et 3 de [20]) assure que, pour chaque θ_j , $n^{-1}[\ell(\theta_j; \mathbf{Y}) - \ell(\theta_0; \mathbf{Y})]$ converge \mathbb{P}_0 -presque sûrement vers un certain $\varepsilon_j < \varepsilon_0 < 0$, le majorant ε_0 étant déterministe. Le choix de $\varepsilon = -\varepsilon_0/2$ permet dès lors de conclure. \square

La Proposition 1 est la combinaison des Propositions 4 et 5.

7.4 Preuve de la Proposition 2

La Proposition 2 est un corollaire de l'inégalité "de mélange" exposée dans la Proposition 6, dans laquelle on reconnaîtra une évolution de la Proposition 3.

Soit K un nombre de régimes fixé. Nous complétons la loi a priori ν_K sur Θ_K en une loi a priori $\nu_K \otimes \nu$ sur $\Theta_K \times \Psi$ avec la caractérisation suivante de ν :

– sous ν , $\{\psi_{(0,0)^T w}\}, \{\psi_{(1,0)^T w}\}, \{\psi_{(0,1)^T w}\}$ sont i.i.d de loi Dirichlet(1/2, 1/2, 1/2).

La nouvelle statistique de mélange d'ordre K est définie par

$$\bar{q}_K(\mathbf{Y}) = \int_{\Theta_K \times \Psi} f(\mathbf{Y}; \theta, \psi) \nu_K(d\theta) \nu(d\psi).$$

C'est à elle que l'on compare la nouvelle log-vraisemblance maximisée dans la proposition suivante.

Proposition 6. Pour tout nombre de régimes $K \geq 1$,

$$0 \leq \sup_{\theta \in \Theta_K} \sup_{\psi \in \Psi} \log \frac{f(\mathbf{Y}; \theta, \psi)}{\bar{q}_K(\mathbf{Y})} \leq \frac{\dim(\Theta_K)}{2} \log nN + \frac{\dim(\Psi)}{2} \log N \\ + \frac{3K}{2\tau^2} \max_{i \leq N, t \leq n} [Y_t^i]^2 + c_{KnN} + \frac{3K}{2} \log \left(\frac{\tau^2}{3K\sigma^2} + \frac{1}{nN} \right) + 3 \log 2. \quad (14)$$

Cette fois, les statistiques $\bar{q}_K(\mathbf{y}|\mathbf{z}, \mathbf{w})$, $\bar{q}_K(\mathbf{z})$ et $\bar{q}_K(\mathbf{w})$ sont définies à partir de (14) en substituant à $f(\mathbf{Y}; \theta, \psi)$ soit $f(\mathbf{y}|\mathbf{z}, \mathbf{w}; \theta, \psi)$, soit $f(\mathbf{z}; \theta)$, soit $f(\mathbf{w}; \psi)$, respectivement. On peut noter que $q_K(\mathbf{z}) = \bar{q}_K(\mathbf{z})$.

Démonstration. On se fonde cette fois sur les deux égalités

$$f(\mathbf{y}; \theta, \psi) = \sum_{\mathbf{w}} \sum_{\mathbf{z}} f(\mathbf{y}|\mathbf{z}, \mathbf{w}; \theta, \psi) f(\mathbf{z}; \theta) f(\mathbf{w}; \psi), \\ \bar{q}_K(\mathbf{y}) = \sum_{\mathbf{w}} \sum_{\mathbf{z}} \bar{q}_K(\mathbf{y}|\mathbf{z}, \mathbf{w}) \bar{q}_K(\mathbf{z}) \bar{q}_K(\mathbf{w})$$

(ici et par la suite les arguments \mathbf{w} et \mathbf{z} parcourent respectivement $\{1, 2, 3\}^N$ et $(\{1, \dots, K\}^n)^N$ avec les conditions initiales $z_1^i = 1$) pour mettre en évidence que

$$\sup_{\theta \in \Theta_K} \sup_{\psi \in \Psi} \log \frac{f(\mathbf{y}; \theta, \psi)}{\bar{q}_K(\mathbf{y})} \leq \sup_{\psi \in \Psi} \max_{\mathbf{w}} \left\{ \log \frac{f(\mathbf{w}; \psi)}{\bar{q}_K(\mathbf{w})} \right\} \\ + \sup_{\theta \in \Theta_K} \sup_{\psi \in \Psi} \max_{\mathbf{w}} \max_{\mathbf{z}} \left\{ \log \frac{f(\mathbf{z}; \theta)}{\bar{q}_K(\mathbf{z})} + \log \frac{f(\mathbf{y}|\mathbf{z}, \mathbf{w}; \theta, \psi)}{\bar{q}_K(\mathbf{y}|\mathbf{z}, \mathbf{w})} \right\}. \quad (15)$$

Une adaptation évidente des définitions de $I^w, T_k^i, N_k^w, \bar{y}_k^w$ introduites en Section 7.3 permet de transposer la démonstration de la Proposition 3 et d'obtenir le contrôle du second terme du membre de droite de (15).

Grâce au choix de l'a priori sur ψ , on démontre aussi que pour tout $\psi \in \Psi$ et \mathbf{w} ,

$$\log \frac{f(\mathbf{w}; \psi)}{\bar{q}(\mathbf{w})} \leq 8(N^{(0,0)} + 1/2)(N^{(1,0)} + 1/2)(N^{(0,1)} + 1/2),$$

où $N^{(0,0)}, N^{(1,0)}, N^{(0,1)}$ sont les nombres respectifs de patients sains, vestibulaires et hémiplésiques. Par convexité, le majorant ci-dessus est plus petit que $3 \log N + 3 \log 2 + 3 \log(1/3 + 1/2N) \leq 3 \log N + 3 \log 2$, ce qui conclut la preuve. \square

La preuve de la Proposition 2 se fait facilement dans le même esprit que celle de la Proposition 1, mais à partir de la Proposition 6 plutôt que de la Proposition 3. Nous omettons ici les détails de cette adaptation aisée.

Remerciements

Ce travail bénéficie des soutiens financiers de l'Université Paris Descartes (via son programme *Bonus Qualité Recherche 2008*) ainsi que du CNRS (via son programme interdisciplinaire *Longévité et vieillissement* pour l'année 2008).

Les auteurs remercient Adélaïde Marquer (Hôpital Fernand-Widal), Céline Quhen (LNRS et Hôpital Lariboisière) pour la partie médicale, Benjamin Favetto (MAP5), Valentine Genon-Catalot (MAP5), Yves Rozenholc (MAP5) et Adeline Samson (MAP5) pour la partie statistique. Ils sont aussi reconnaissants envers les rapporteurs pour des remarques très pertinentes sur la première version du manuscrit.

k	1	2	3	4	5	6	7	8	9	10
$\hat{\ell}_1(k)$	-20964.45	-15095.46	-14302.02	-14118.07	-13895.69	-13790.21	-13787.56	-13731.94	-13695.73	-13686.34
BIC	-20977.56	-15130.42	-14367.58	-14222.96	-14048.66	-14000.00	-14062.91	-14081.59	-14128.42	-14210.82
BIC cumulé	-20996.42	-15190.43	-14499.59	-14466.54	-14451.02	-14616.52	-14959.15	-15331.86	-15815.76	-16427.00

TABLEAU 1 – Valeurs des log-vraisemblances maximisées pour le premier modèle (notée $\hat{\ell}_1(k)$ pour le modèle d'ordre k) et leur versions pénalisées par les critères BIC et “BIC cumulé”.

k	1	2	3	4	5	6	7	8	9	10
$\hat{\ell}_2(k)$	-16154.30	-14806.19	-14182.05	-14095.74	-13876.33	-13776.84	-13760.86	-13715.88	-13680.90	-13664.92
BIC	-16176.82	-14850.56	-14257.02	-14210.04	-14038.71	-13996.04	-14045.62	-14074.94	-14123.00	-14198.80
BIC cumulé	-16197.76	-14924.13	-14414.09	-14490.16	-14489.09	-14672.07	-15012.86	-15407.69	-15904.30	-16520.44

TABLEAU 2 – Valeurs des log-vraisemblances maximisées pour le second modèle (notée $\hat{\ell}_2(k)$ pour le modèle d'ordre k) et leur versions pénalisées par les critères BIC et “BIC cumulé”.

$$\{\hat{P}_{k\ell}\} = \begin{pmatrix} 0.917 & 0.030 & 0.053 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.111 & 0.000 & 0.889 \\ 0.204 & 0.000 & 0.089 & 0.243 & 0.463 \\ 0.009 & 0.000 & 0.045 & 0.920 & 0.026 \\ 0.017 & 0.752 & 0.090 & 0.092 & 0.049 \end{pmatrix}$$

k	1	2	3	4	5
\hat{m}_k	-4.058 ± 0.090	-2.837 ± 0.189	-8.725 ± 0.182	-1.639 ± 0.979	-2.703 ± 0.185
$\hat{m}_k + (1, 0)\hat{\beta}_k$	0.404 ± 0.223	-3.386 ± 0.517	-7.917 ± 0.404	-0.959 ± 0.206	-3.019 ± 0.457
$\hat{m}_k + (0, 1)\hat{\beta}_k$	0.314 ± 0.204	-2.711 ± 0.439	-7.301 ± 0.391	-1.405 ± 0.218	-3.106 ± 0.403

TABLEAU 3 – Résultats de l’ajustement du premier modèle d’ordre 5. Les coefficients sont arrondis au millième. Les écarts-types apparaissent après le symbole ±.

$$\{\hat{P}_{k\ell}\} = \begin{pmatrix} 0.860 & 0.039 & 0.101 \\ 0.026 & 0.929 & 0.045 \\ 0.556 & 0.345 & 0.099 \end{pmatrix}$$

k	1	2	3
\hat{m}_k	-3.764 ± 0.078	-1.925 ± 0.065	-8.672 ± 0.179
$\hat{m}_k + (1, 0)\hat{\beta}_k$	-2.833 ± 0.168	-0.734 ± 0.138	-7.687 ± 0.375
$\hat{m}_k + (0, 1)\hat{\beta}_k$	-2.428 ± 0.190	0.741 ± 0.152	-7.040 ± 0.405
w	1	2	3
$\psi_{(0,0)^T w}$	1	0.000	0.000
$\psi_{(1,0)^T w}$	0	0.750	0.250
$\psi_{(0,1)^T w}$	0	0.667	0.333

TABLEAU 4 – Résultats de l'ajustement du second modèle d'ordre 3. Les coefficients sont arrondis au millièème. Les écarts-types apparaissent après le symbole \pm .

Références

- [1] P. S. Albert, H. F. McFarland, M. E. Smith, and Frank . A. Time series for modelling counts from a relapsed-remitting disease : application to modelling disease activity in multiple sclerosis. *Statistics in Medicine*, (24) :453–466, 1994.
- [2] R. MacKay Altman. Mixed hidden Markov models : an extension of the hidden Markov model to the longitudinal data setting. *J. Amer. Statist. Assoc.*, 102(477) :201–210, 2007.
- [3] J-M. Bardet and P. Bertrand. Identification of the multiscale fractional Brownian motion with biomechanical applications. *J. Time Ser. Anal.*, 28(1) :1–52, 2007.
- [4] P. Bertrand, J-M. Bardet, M. Dabonneville, A. Mouzat, and P. Vaslin. Automatic determination of the different control mechanisms in upright position by a wavelet method. *IEEE Engineering in Medicine and Biology Society*, 2 :1163–1166, 2001.
- [5] C. Biernacki, G. Celeux, and G. Govaert. Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(7) :719–725, 2000.
- [6] O. Cappé, E. Moulines, and T. Rydén. *Inference in hidden Markov models*. Springer Series in Statistics. Springer, New York, 2005.
- [7] A. Chambaz, A. Garivier, and E. Gassiat. A minimum description length approach to hidden Markov models with Poisson and Gaussian emissions. Application to order identification. *J. Statist. Plann. Inference*, pages 1–16, 2008.
- [8] L. Chiari, A. Cappello, D. Lenzi, and U. Della Croce. An improved technique for the extraction of stochastic parameters from stabilograms. *Gait Posture*, 12(3) :225–234, 2000.
- [9] J.-J. Collins and C. J. De Lucas. Open-loop and closed-loop control of posture : a random-walk analysis of center-of-pressure trajectories. *Exp. Brain Res.*, 1993.
- [10] Lee D. Davisson, Robert J. McEliece, Michael B. Pursley, and Mark S. Wallace. Efficient universal noiseless source codes. *IEEE Trans. Inform. Theory*, 27(3) :269–279, 1981.
- [11] L. Finesso. Consistent estimation of the order for Markov and hidden Markov chains. *PhD Thesis, University of Maryland*, 1991.
- [12] T. D. Frank, A. Daffertshofer, and P. J. Beek. Multivariate ornstein-uhlenbeck processes with mean-field dependent coefficients : Application to postural sway. *Phys. Rev. E*, 63(1) :011905, Dec 2000.
- [13] P. A. Fransson, A. Hafström, M. Karlberg, M. Magnusson, A. Tjäder, and R. Johansson. Postural control adaptation during galvanic vestibular and vibratory proprioceptive stimulation. *IEEE Trans. Biomed. Eng.*, 50(12) :1310–1319, 2003.
- [14] P. A. Fransson, R. Johansson, A. Hafström, and M. Magnusson. Methods for evaluation of postural control adaptation. *Gait Posture*, 12(1) :14–24, 2000.
- [15] P. A. Fransson, E. K. Kristinsdottir, A. Hafström, M. Magnusson, and R. Johansson. Balance control and adaptation during vibratory perturbations in middle-aged and elderly humans. *Eur. J. Appl. Physiol.*, 91(5-6) :595–603, 2004.
- [16] E. Gassiat. Likelihood ratio inequalities with applications to various mixtures. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(6) :897–906, 2002. En l’honneur de J. Bretagnolle, D. Dacunha-Castelle, I. Ibragimov.

- [17] T. Koski. *Hidden Markov models for bioinformatics*, volume 2 of *Computational Biology Series*. Kluwer Academic Publishers, Dordrecht, 2001.
- [18] J. Laurens and J. Droulez. Bayesian processing of vestibular information. *Biol. Cybern.*, 96(4) :389–404, 2007.
- [19] J. C. Lepecq, C. De Waele, S. Mertz-Josse, C. Teyssède, P. T. Huy, P. M. Baudonnière, and P-P. Vidal. Galvanic vestibular stimulation modifies vection paths in healthy subjects. *J. Neurophysiol.*, 95(5) :3199–3207, 2006.
- [20] B. G. Leroux. Maximum-likelihood estimation for hidden Markov models. *Stochastic Process. Appl.*, 40(1) :127–143, 1992.
- [21] S. E. Levison, L. R. Rabiner, and M. M. Sondhi. An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition. *Bell System Tech. J.*, 62 :1035–1074, 1983.
- [22] C. C. Liu and P. Narayan. Order estimation and sequential universal data compression of a hidden markov source by the method of mixtures. *Canad. J. Statist.*, 30(4) :573–589, 1994.
- [23] T. Mergner, G. Schweigart, C. Maurer, and A. Blümle. Human postural responses to motion of real and virtual visual environments under different support base conditions. *Exp. Brain Res.*, 167(4) :535–556, 2005.
- [24] J. C. Nash. *Compact numerical methods for computers : linear algebra and function minimisation*. John Wiley & Sons Inc., New York, 1979. A Halsted Press Book.
- [25] K. M. Newell, S. M. Slobounov, E. S. Slobounova, and P. C. Molenaar. Stochastic processes in postural center-of-pressure profiles. *Exp. Brain Res.*, 113(1) :158–164, 1997.
- [26] R. A. Olshen, E. N. Bide, Marilyn P. Wyatt, and D. H. Sutherland. Gait analysis and the bootstrap. *Ann. Statist.*, 17(4) :1419–1440, 1989.
- [27] R Development Core Team. *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0.
- [28] A.M. Sabatini. A statistical mechanical analysis of postural sway using non-gaussian farima stochastic models. *IEEE Trans. Biomed. Eng.*, 47(9) :1219–1227, 2000.