

Processus stochastiques pour la classification selon le style postural

Christophe Denis

Université Paris Descartes
MAP5, UMR CNRS 8145

En collaboration avec

A. Chambaz (MAP5, UMR CNRS 8145)

A. Samson (MAP5, UMR CNRS 8145)

et P-P. Vidal (CESEM, UMR CNRS 8194)

17/04/2011

Colloque Jeunes probabilistes et statisticiens

Introduction

Modélisation

Procédure de classification

Application aux données réelles

Perspectives

Introduction

Modélisation

Procédure de classification

Application aux données réelles

Perspectives

Contexte

- Le maintien postural est le fruit du traitement dynamique, par le cerveau, d'informations provenant des systèmes :
 - ▶ visuel,
 - ▶ vestibulaire,
 - ▶ proprioceptif.
- Une stratégie optimale de maintien postural repose sur l'exploitation des trois types d'informations.
- A contrario, privilégier un unique type d'information (typiquement visuelle) est une stratégie ne permettant pas de gérer l'imprévu.
- Les objectifs de l'étude du maintien postural sont de permettre :
 - ▶ à un médecin généraliste d'identifier chez un patient, de manière simple, un trouble de la posture ;
 - ▶ la mise au point de programme de physiothérapie adapté au patient.

Notre jeu de données

- 54 patients : 32 patients sains, 22 hémiplésiques.
- Pour chaque patient :
 - ▶ covariables W_i (âge, genre, poids, taille, latéralité)
 - ▶ mesures $(X_{1:N}) = (X_1, \dots, X_N)$ prises lors de différents protocoles.
- Pour chaque protocole, le centre de pression maximale exercée de chaque pied a été enregistré au cours du temps sur une plateforme de force :
 - ▶ $X_i = (L_i, R_i)$ où $L_i = (L_i^1, L_i^2) \in \mathbb{R}^2$ et $R_i = (R_i^1, R_i^2) \in \mathbb{R}^2$.
- deux protocoles considérés :

protocole	0→15s	15→50s	50→70s
1	pas de perturbation	stimulation musculaire	pas de perturbation
2		yeux fermés stimulation musculaire	

Objectifs principaux de l'étude

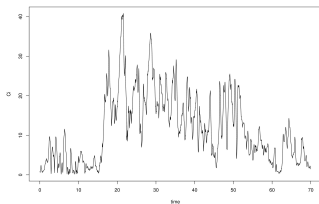
- Classer les patients en terme de maintien postural.
- Construction de deux classifieurs :
 - ▶ ϕ_1 fonction des covariables et d'une trajectoire associée à un protocole
 - ▶ ϕ_2 fonction des covariables et des trajectoires associées aux deux protocoles

Transformation des données brutes

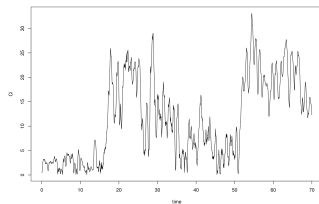
- Les mesures $(X_{1:N})$ sont de très grande dimension.
- On propose de substituer à $(X_{1:N})$ une quantité Y de faible dimension.
- Première étape : réduire $(X_{1:N})$ à $(C_{1:N})$, avec $C_i \in \mathbb{R}$
 1. pour tout i , soit $B_i = (B_i^1, B_i^2)$
 - B_i^1 barycentre de (L_i^1, R_i^1)
 - B_i^2 barycentre de (L_i^2, R_i^2)
 2. choix d'un point de référence b ,
 3. définition de

$$C_i = \|B_i - b\|_2.$$

Trajectoire $i \mapsto C_i$



stim. musculaire et visuelle



stim. musculaire

Introduction

Modélisation

Procédure de classification

Application aux données réelles

Perspectives

Modélisation de $C_{1:N}$ (1/3)

- Trajectoire $C_{1:N}$ considérée comme l'observation d'un processus stochastique $(C(t))_{[T_0, T]}$ défini par :

$$\begin{cases} dC(t) = b(C(t), \phi)dt + a(C(t), \sigma)dW(t) \\ C(T_0) = C_0, \end{cases}$$

- ▶ $W(t)$ processus de Wiener standard,
- ▶ b le drift et a la volatilité sont des fonctions paramétriques connues.
- ▶ ϕ et σ paramètres inconnus.
- ▶ $T_0 = 0.025s$ et $T = 70s$.

Modélisation de $C_{1:N}$ (2/3)

- Phase de perturbation d'un protocole implique le changement des paramètres de $(C(t))_{t \in [T_0, T]}$:
 - ▶ au voisinage du début et de la fin de la phase de perturbation
 - ▶ éventuellement aussi pendant.

- On note la suite des instants de ruptures,

$$T_0 = \delta < T_1 = \tau_1 \delta < \dots < T_{k-1} = \tau_{k-1} \delta < T_K = T$$

- ▶ avec $K = 3$ ou 4 correspondant à 2 ou 3 instants de ruptures.
- Sur $[T_{k-1}, T_k]$:

$$\begin{cases} dC(t) = b(C(t), \phi_k)dt + a(C(t), \sigma_k)dW(t) \\ C(T_{k-1}) = C_{\tau_{k-1}}, \end{cases}$$

- Dernière étape : modélisation des formes paramétrique a et b .

Modélisation de $C_{1:N}$ (3/3)

- On modélise $C(t)$ par un processus de Cox-Ingersoll-Ross (CIR).
- Modèle : pour tout $k \in \{1, \dots, K\}$, $(C(t))_{t \in [T_{k-1}, T_k]}$ est définie par :

$$\begin{cases} dC(t) = \lambda_k(\mu_k - C(t))dt + \sigma_k \sqrt{C(t)}dW(t) \\ C(T_{k-1}) = C_{T_{k-1}}, \end{cases}$$

- ▶ avec $C_{T_{k-1}}$ observation au temps T_{k-1} ,
- ▶ λ_k est l'échelle de temps du processus,
- ▶ μ_k est la moyenne de la distribution asymptotique,
- ▶ σ_k^2 est le paramètre de variance du processus.
- Propriétés élémentaires de $(C(t))_{t \in [T_{k-1}, T_k]}$:
 - ▶ Si $\frac{2\mu_k \lambda_k}{\sigma_k^2} \geq 1$ alors le processus reste positif et admet une distribution stationnaire,
 - ▶ loi Gamma de paramètres $\frac{2\mu_k \lambda_k}{\sigma_k^2}$ et $\frac{\sigma_k^2}{2\lambda_k}$.

Procédure d'estimation des paramètres du processus $C(t)$.

- On note $\theta_k = (\lambda_k, \mu_k, \sigma_k)$.
- Estimation des paramètres en deux étapes :
 - ▶ estimation de la suite des instants de ruptures $\Rightarrow \{\hat{T}_1, \dots, \hat{T}_{K-1}\}$.
 - ▶ pour tout $k \in \{1, \dots, K\}$, estimation du paramètre θ_k de $(C(t))_{[\hat{T}_{k-1}, \hat{T}_k]}$.
- Estimation de la suites des instants de ruptures fondée sur la minimisation d'un contraste empirique [Lavielle (2005)].
- Estimation des paramètres du processus fondée sur la minimisation d'un contraste empirique [Kessler (1997)].

Estimation des instants de rupture

- Contraste proposé par Lavielle (2005) :

$$\mathcal{J}(\tau, C_{1:N}) = \frac{1}{N} \sum_{k=1}^K (\tau_k - \tau_{k-1} + 1) \log(\gamma_{\tau_{k-1}:\tau_k}^2),$$

- ▶ où $\gamma_{\tau_{k-1}:\tau_k}^2$ estimateur empirique de la variance de $C_{\tau_{k-1}:\tau_k}$.
- l'estimateur $\hat{\tau}_N$ de la suite des instants de rupture est définie par :

$$\hat{\tau}_N = \arg \min_{\tau} \mathcal{J}(\tau, C_{1:N}).$$

- Sous des des hypothèses générales, pour tout $k \in \{1, \dots, K-1\}$:

$$P(|\hat{\tau}_{N,k} - \tau_k| > \eta) \rightarrow 0, \text{ quand } \eta \rightarrow \infty \text{ et } N \rightarrow \infty.$$

Estimation des paramètres du processus

- Pour tout $k \in \{1, \dots, K-1\}$, l'approximation discrète par le schéma d'Euler Marumaya de $(C(t))_{[\hat{T}_{k-1}, \hat{T}_k]}$ est définie, $i \in \{\hat{T}_{k-1}, \dots, \hat{T}_k\}$ par :

$$\tilde{C}_{i+1} = (1 - \delta\lambda_k)\tilde{C}_i + \delta\lambda_k\mu_k + \sigma_k\sqrt{\delta}\sqrt{\tilde{C}_i}\eta_{i+1},$$

- Le contraste $\mathcal{L}(\theta_k)$ proposé par Kessler (1997) :

$$\mathcal{L}_k(\theta_k) = \sum_{i=\hat{T}_{k-1}}^{\hat{T}_k-1} \frac{(C_{i+1} - \theta_{1,k}C_i - \theta_{2,k})^2}{C_i\theta_{3,k}^2} + (\hat{T}_k - \hat{T}_{k-1} + 1) \log(\theta_{3,k}^2),$$

avec $\theta_{1,k} = 1 - \lambda_k\delta$, $\theta_{2,k} = \delta\lambda_k\mu_k$ et $\theta_{3,k}^2 = \delta\sigma_k^2$.

- On a $\hat{\theta}_k = \arg \min_{\theta_k \in \mathbb{R}_+^3} \mathcal{L}_k(\theta_k)$
- Sous des hypothèses générales on peut montrer la consistance des estimateurs $\hat{\theta}_k$.

Estimation des instants de ruptures sur les données réelles

Sujets normaux

protocole	$K = 3$		$K = 4$		
1	18.28 (4.37)	50.66 (10.53)	16.62 (5.02)	31.90 (13.75)	52.97 (6.85)
2	19.75 (7.83)	52.62 (4.87)	15.84 (4.32)	31.59 (12.47)	52.68 (4.80)

Sujets hémiplésiques

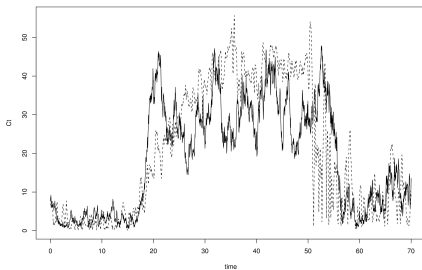
protocole	$K = 3$		$K = 4$		
1	15.27 (3.13)	45.27 (13.78)	14.27 (3.89)	31.18 (11.13)	51.86 (8.84)
2	16.95 (5.39)	48.13 (11.60)	13.95 (7.15)	30.45 (14.38)	55.36 (3.35)

Estimation des paramètres du processus sur les données réelles

protocole	param.	Sujets normaux		
		$k = 1$	$k = 2$	$k = 3$
1	$\theta_{1,k}$	0.98 (0.01)	0.99 (0.01)	0.98 (0.02)
	$\theta_{2,k}$	0.06 (0.04)	0.14 (0.12)	0.11 (0.08)
	$\theta_{3,k}^2$	0.04 (0.03)	0.03 (0.04)	0.03 (0.04)
2	$\theta_{1,k}$	0.98 (0.01)	0.99 (0.01)	0.98 (0.02)
	$\theta_{2,k}$	0.09 (0.08)	0.22 (0.26)	0.15 (0.14)
	$\theta_{3,k}^2$	0.06 (0.06)	0.05 (0.04)	0.05 (0.06)

protocole	param.	Sujets hémiplégiques		
		$k = 1$	$k = 2$	$k = 3$
1	$\theta_{1,k}$	0.96 (0.02)	0.98 (0.01)	0.98 (0.02)
	$\theta_{2,k}$	0.13 (0.09)	0.27 (0.30)	0.16 (0.13)
	$\theta_{3,k}^2$	0.08 (0.05)	0.05 (0.04)	0.05 (0.05)
2	$\theta_{1,k}$	0.96 (0.02)	0.98 (0.02)	0.97 (0.02)
	$\theta_{2,k}$	0.15 (0.10)	0.37 (0.36)	0.36 (0.62)
	$\theta_{3,k}^2$	0.1 (0.06)	0.1 (0.10)	0.09 (0.08)

Exemple de trajectoire



Trajectoire $C_{1:N}$ (trait plein)
Trajectoire simulée (pointillés)

Introduction

Modélisation

Procédure de classification

Application aux données réelles

Perspectives

Modèle statistique pour la classification

- Construction des nouvelles mesures résumées fondée sur l'estimation des temps de rupture et des paramètres du processus.
- Plus formellement, à une trajectoire $C_{1:N}$ on associe la mesure résumée Y_K définie pour $K \in \{3, 4\}$ par :

$$Y_K = (\hat{\mu}_1, \dots, \hat{\mu}_K, \hat{\sigma}_1^2, \dots, \hat{\sigma}_K^2, \hat{T}_1, \dots, \hat{T}_{K-1}),$$

- On note Y_K^j la mesure résumée associée au protocole j .
- Observation :

$$O = (W, Y_3^1, Y_4^1, Y_3^2, Y_4^2, A) \sim P_0.$$

Définition des classifieurs

- On note $Z = (W, Y_3^1, Y_4^1, Y_3^2, Y_4^2)$ et \mathcal{S} l'ensemble des classifieurs fonction de Z .
- Pour tout $j \in \{1, 2\}$, $K \in \{3, 4\}$ on définit $S_K^j \in \mathcal{S}$:

$$S_K^j(W, Y_K^j) = \mathbb{1}\{\eta_K^j(W, Y_K^j) \geq \frac{1}{2}\},$$

- avec $\eta_K^j(W, Y_K^j) = P_0(A = 1 | W, Y_K^j)$;
 - le classifieur S_K^j est le classifieur de Bayes basé sur W et Y_K^j .
- Pour tout $(K_1, K_2) \in \{3, 4\}$, on définit $S_{K_1, K_2} \in \mathcal{S}$:

$$S_{K_1, K_2}(W, Y_{K_1}^1, Y_{K_2}^2) = \mathbb{1}\{\eta_{K_1, K_2}(W, Y_{K_1}^1, Y_{K_2}^2) \geq \frac{1}{2}\},$$

- $\eta_{K_1, K_2}(W, Y_{K_1}^1, Y_{K_2}^2) = P_0(A = 1 | W, Y_{K_1}^1, Y_{K_2}^2)$
 - le classifieur S_{K_1, K_2} est le classifieur de Bayes basé sur W et $(Y_{K_1}^1, Y_{K_2}^2)$.

Procédure de classification

- On définit pour $s \in S$, $R^{(P_0)}(s) = P_0(s(Z) \neq A)$.
- On définit les classifieurs ϕ_1 et ϕ_2 comme :

$$\phi_1 \in \arg \min_{S_K^j} R^{(P_0)}(S_K^j)$$

$$\phi_2 \in \arg \min_{S_{K_1, K_2}} R^{(P_0)}(S_{K_1, K_2}).$$

- Objectif : construire $\hat{\phi}_1$ et $\hat{\phi}_2$ estimateur de ϕ_1 et ϕ_2 .
- Procédure de classification en deux étapes :
 - ▶ construction des ensembles \mathcal{H}_1 et \mathcal{H}_2 d'estimateurs candidats pour ϕ_1 et ϕ_2 :
 - $\mathcal{H}_1 = \{\hat{S}_3^1, \hat{S}_4^1, \hat{S}_3^2, \hat{S}_4^2\}$
 - $\mathcal{H}_2 = \{\hat{S}_{3,3}, \hat{S}_{3,4}, \hat{S}_{4,3}, \hat{S}_{4,4}\}$
 - ▶ sélection de $\hat{\phi}_1$ (resp. $\hat{\phi}_2$) parmi l'ensemble \mathcal{H}_1 (resp. \mathcal{H}_2) fondée sur le principe de validation croisée.

Introduction

Modélisation

Procédure de classification

Application aux données réelles

Perspectives

Application de la procédure de classification aux données réelles

- On évalue les performances de la méthode de classification par la règle du leave-one-out :

	$\hat{\phi}_1$	$\hat{\phi}_2$
% bien classé	83 %	80 %

- Résultats de l'extension de la procédure de classification par ajout de mesures résumées proposées [Chambaz et Denis (2012)] :

	$\hat{\phi}_1$	$\hat{\phi}_2$
% bien classé	83 %	91 %

Introduction

Modélisation

Procédure de classification

Application aux données réelles

Perspectives

Conclusions et perspectives

- Conclusion :
 - ▶ modélisation par un processus de diffusion apporte de l'information \Rightarrow étape de modélisation cruciale.
 - ▶ résultats obtenus très satisfaisants 91%
- Perspectives :
 - ▶ prise en compte des informations sur l'orientation des patients durant les protocoles \Rightarrow modélisation de $B_{1:N}$ par un processus bi-dimensionnel
 - ▶ étendre la procédure de classification où A prend plus que deux valeurs