Bruit à dépendance longue dans les Chaînes de Markov Cachées (cas gaussien)

Stéphane Derrode
Séminaire PHARE'09, ENSI, Tunisie

École Centrale Marseille & Institut Fresnel (CNRS)

28 octobre 2009

Modélisation

Modélisation stochastique des données

- Processus des observations :
  $$\boldsymbol{Y} = \{Y_1, \ldots, Y_N\}, \quad Y_n \in \mathbb{R}.$$
- Processus (caché) des classes :
  $$\boldsymbol{X} = \{X_1, \ldots, X_N\}, \quad X_n \in \Omega = \{1, \ldots, K\}.$$

➥ $\boldsymbol{X}$ est une chaîne de Markov :

$$\boxed{p\left(X_{n+1} \mid X_1, \ldots, X_n\right) = p\left(X_{n+1} \mid X_n\right)}.$$

Elle est cachée et doit être estimée à partir de $\boldsymbol{Y} = \boldsymbol{y}$.

Hypothèses

[H1] $\boldsymbol{X}$ est une chaîne de Markov stationnaire

- Loi initiale : $\pi_j = p(X_1 = j)$
- Loi stationnaire (matrice stochastique $\mathbf{A}$) :

$$a_{j|i} = p(X_2 = j \,|X_1 = i)$$

[H2] Les V.A. Y_n sont indépendantes conditionnellement à $\boldsymbol{X}$

$$p(\boldsymbol{y} \,|\boldsymbol{x}) = \prod_{n=1}^{N} p(y_n \,|\boldsymbol{x})$$

[H3] $p(y_n \,|\boldsymbol{x}) = p(y_n \,|x_n) = f_{x_n}(y_n)$ (loi d'attache aux données)

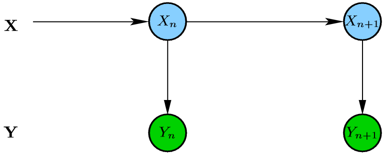
Structure de dépendance

Globalement

Loi *a priori* $$p(\boldsymbol{x}) = p(x_1) \prod_{n=1}^{N-1} p(x_{n+1} | x_n)$$

Loi du couple $$p(\boldsymbol{x}, \boldsymbol{y}) = p(x_1)\ f_{x_1}(y_1) \prod_{n=1}^{N-1} p(x_{n+1} | x_n)\ f_{x_{n+1}}(y_{n+1}).$$

Localement (contextuellement)



$$p(x_{n+1}, y_{n+1} | x_n, y_n) = p(x_{n+1} | x_n)\ f_{x_{n+1}}(y_{n+1})$$

Processus à dépendance longue

Soit $\boldsymbol{Y} = (Y_n)_{n\in\mathbb{Z}}$, un processus stationnaire au second ordre, i.e.

$$\boxed{\gamma(k) = E\left[Y_n \; Y_{n-k}\right] - E\left[y_n\right] \; E\left[y_{n-k}\right]}$$

ne dépend pas de n. $(\gamma(k))_{k\in\mathbb{Z}}$ est appelée « famille de covariances ».

Définition

Le processus est à dépendance longue si $\exists\alpha \in]0,1]$, $C \in \mathbb{R}$ tels que :

$$\lim_{k \leftarrow +\infty} k^{\alpha} \; \gamma(k) = C.$$

On note $\gamma(k) \sim_{+\infty} C \; k^{-\alpha}$. Lorsque $\alpha > 1$, on dit que $\boldsymbol{Y}$ est à dépendance intermédiaire.

➥ Exemple : Processus 2-stationnaire. C'est un processus de moyenne m et de famille de covariance

$$\boxed{\forall k \in \mathbb{Z}, \quad \gamma(k) = \sigma^2 \; \left(|k| + 1\right)^{-\alpha}.}$$

Autres exemples

➥ Bruit gaussien fractionnaire : c'est un processus auto-similaire à accroissement stationnaire $Y_n = Z_{n+1} - Z_n + m$, tel que le processus $\boldsymbol{Z}$ est une mouvement brownien fractionnaire. Famille de covariance :

$$\boxed{\forall k \in \mathbb{Z}, \quad \gamma(k) = \frac{\sigma^2}{2}\left[|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}\right]},$$

avec $\sigma^2 = E\left[Z_0\right]$. Si $H > 1/2$, alors $\boldsymbol{Y}$ est à dépendance longue.

➥ Processus FARIMA : On peut construire un FARIMA de paramètre $d \in [-1/2, 1/2] - \{0\}$ en filtrant un processus bruit blanc $\boldsymbol{Z}$ de variance σ^2 de la manière suivante : $Y = a \star Z + m$, avec $a_k = \Gamma(d+k) / \left(\Gamma(d)\,\Gamma(k+1)\right)$. Famille de covariance :

$$\boxed{\forall k \in \mathbb{Z}, \quad \gamma(k) = \sigma^2\,\frac{\Gamma(k+d)\,\Gamma(1-2d)}{\Gamma(k-d+1)\,\Gamma(d)\,\Gamma(1-d)}}.$$

Si $d > 0$, alors $\boldsymbol{Y}$ est à dépendance longue.

Modélisation

➥ $\boldsymbol{Z} = (\boldsymbol{X}, \boldsymbol{Y})$ est une chaîne couple de Markov :

$$\boxed{p\left(z_{n+1} \left| z_{1:n}\right.\right) = p\left(z_{n+1} \left| z_n\right.\right)}.$$

.

Loi du couple $\quad p(\boldsymbol{z}) = p(z_1) \displaystyle\prod_{n=1}^{N-1} p\left(z_{n+1} \left| z_n\right.\right).$

➥ $\boldsymbol{Z} = (\boldsymbol{X}, \boldsymbol{Y})$ est une chaîne couple *partiellement de Markov* :

$$\boxed{p\left(z_{n+1} \left| z_{1:n}\right.\right) = p\left(z_{n+1} \left| x_n, y_{1:n}\right.\right)}.$$

.

Loi du couple $\quad p(\boldsymbol{z}) = p(z_1) \displaystyle\prod_{n=1}^{N-1} p\left(z_{n+1} \left| x_n, y_{1:n}\right.\right).$

Structure de dépendance

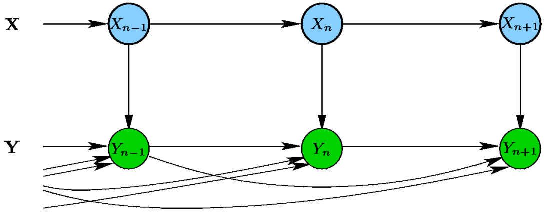
Hypothèses simplificatrices

$$\boxed{p\left(z_{n+1} \,|x_n, y_{1:n}\right) = p\left(y_{n+1} \,|x_n, x_{n+1}, y_{1:n}\right)\; p\left(x_{n+1} \,|x_n, y_{1:n}\right)}$$

[H1] $p\left(x_{n+1} \,|x_n, y_{1:n}\right) = p\left(x_{n+1} \,|x_n\right)$

[H2] $p\left(y_{n+1} \,|x_{n+1}, x_n, y_{1:n}\right) = p\left(y_{n+1} \,|x_{n+1}, y_{1:n}\right)$

Localement (contextuellement)



Simulation d'une CMC-ML

Soit π_k le vecteur de probabilités initiale et $\boldsymbol{A}$ la matrice de transition caractérisant la chaîne de Markov stationnaire. Soit $f_k(.)$ les lois d'attache aux données gaussiennes de moyennes m_k et de variance σ_k^2.

Simulation de $\boldsymbol{X}$

- x_1 est simulé en effectuant un tirage sur $p\left(x_1\right)$.
- $x_n,\ n > 1$ est simulé en effectuant un tirage sur le vecteur de probabilités $p\left(x_n \left| x_{n-1}\right.\right)$.

Simulation de $\boldsymbol{Y}$

- $y_n,\ n \geq 1$ est simulé en effectuant un tirage sur la gaussienne de paramètre $m_{x_{\boldsymbol{n}}}$ et $\sigma^2_{x_{\boldsymbol{n}}}$.

Simulation de $\boldsymbol{Y}$: paramètres des gaussiennes

- $n = 1$ $p\left(y_1 \middle| x_1\right)$ est une gaussienne de paramètres m_{x_1} et $\gamma_{x_1}(0)$.
- $n > 1$ $p\left(y_{n+1} \middle| x_{n+1}, y_{1:n}\right)$ est une gaussienne de paramètres

$$\begin{aligned}
\tilde{m}_{x_{n+1}} &= m_{x_{n+1}} + \Gamma^{2,1}_{x_{n+1}} \left(\Gamma^{n}_{x_{n+1}}\right)^{-1} \left(y_{1:n} - m^{n}_{x_{n+1}}\right) \\
\tilde{\gamma}_{x_{n+1}} &= \gamma_{x_{n+1}}(0) + \Gamma^{2,1}_{x_{n+1}} \left(\Gamma^{n}_{x_{n+1}}\right)^{-1} \Gamma^{1,2}_{x_{n+1}}
\end{aligned}$$

avec

$$m^{n}_{x_{n+1}} = (\underbrace{m_{x_{n+1}}, \ldots, m_{x_{n+1}}}_{n \text{ fois}})$$

$$\Gamma^{2,1}_{x_{n+1}} = \left(\gamma_{x_{n+1}}(n), \ldots, \gamma_{x_{n+1}}(1)\right)$$

$$\Gamma^{n}_{x_{n+1}} = \begin{pmatrix}
\gamma_{x_{n+1}}(0) & \gamma_{x_{n+1}}(1) & \cdots & \gamma_{x_{n+1}}(n-1) \\
\gamma_{x_{n+1}}(1) & \gamma_{x_{n+1}}(0) & \cdots & \gamma_{x_{n+1}}(n-2) \\
\vdots & \vdots & \vdots & \vdots \\
\gamma_{x_{n+1}}(n-1) & \gamma_{x_{n+1}}(n-2) & \cdots & \gamma_{x_{n+1}}(0)
\end{pmatrix}$$

Exemple de simulations ($K = 2$)

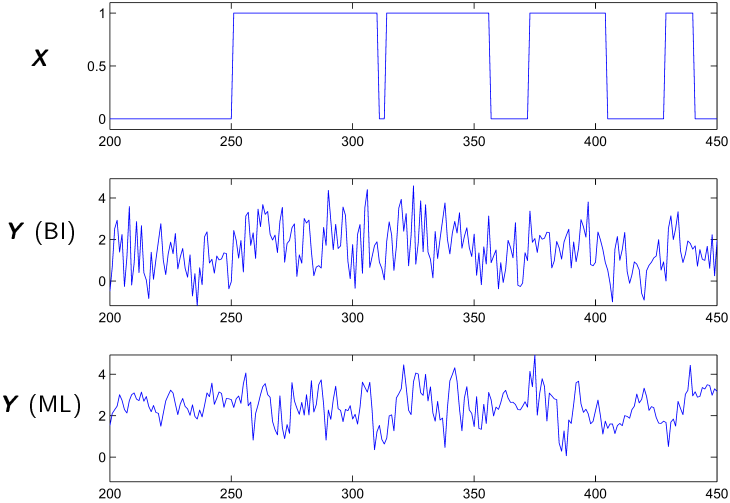
Paramètres de la chaîne de Markov

$$\boldsymbol{\pi} = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}, \quad \boldsymbol{A} = \begin{pmatrix} 0.96 & 0.04 \\ 0.04 & 0.96 \end{pmatrix}$$

Paramètres des lois d'attache aux données (gaussiennes)

- Loi de la classe #1 : $m_1 = 1$, $\sigma_1^2 = 1$
- Loi de la classe #2 : $m_2 = 2$, $\sigma_2^2 = 1$
- Pour la longue dépendance, nous utilisons la famille de covariance $\gamma(k) = \sigma^2 \; (|k| + 1)^{-\alpha}$, avec $\alpha_1 = 0.1$ et $\alpha_2 = 0.9$.

Résultat de simulations



Algorithmes de classification

Décision bayésienne

$$\hat{\mathbf{x}}^{OPT}(\mathbf{y}) = \arg\min_{\hat{\mathbf{x}} \in \Omega^{\mathbf{N}}} E\left[L(\mathbf{x}, \hat{\mathbf{x}}) \,|\, \mathbf{Y} = \mathbf{y}\right]$$

- MAP - Fonction de perte : $L_2(\mathbf{x}, \hat{\mathbf{x}}) = 1_{\{\mathbf{x} \neq \hat{\mathbf{x}}\}}$

$$\hat{\mathbf{x}}^{MAP}(\mathbf{y}) = \arg\max_{\mathbf{x} \in \Omega^{\mathbf{N}}} p(\mathbf{X} = \mathbf{x} \,|\, \mathbf{Y} = \mathbf{y})$$

- MPM - Fonction de perte : $L_1(\mathbf{x}, \hat{\mathbf{x}}) = \sum_{n=1}^{N} 1_{x_{\mathbf{n}} \neq \hat{x}_{\mathbf{n}}}$

$$\hat{x}_n^{MPM}(\mathbf{y}) = \arg\max_{x_n \in \Omega} p(X_n = x_n \,|\, \mathbf{Y} = \mathbf{y}) = \alpha_n(k)\, \beta_n(k)$$

↪ Tous basés sur l'algorithme "*forward/backward*" de Baum (1970).

Algo. « *forward/backward* »

Probabilités « *forward* »

$$\boxed{\alpha_n(x_n) = p\left(x_n \,|y_{1:n}\right) = \frac{1}{S_n}\, p\left(x_n, y_n \,|y_{1:n-1}\right)},$$

avec $S_n = p\left(y_n \,|y_{1:n-1}\right)$, $n > 1$ le facteur de normalisation.

- $n = 1$:

$$\alpha_1(x_1) = p\left(x_1 \,|y_1\right).$$

- $1 \leq n < N - 1$:

$$\alpha_{n+1}(x_{n+1}) = \frac{1}{S_{n+1}} \sum_{x_n} \alpha_n(x_n)\, p\left(z_{n+1} \,|x_n, y_{1:n}\right).$$

Probabilités « *backward* »

$$\boxed{\beta_n(x_n) = \frac{p\left(y_{n+1:N} \,|x_n\right)}{p\left(y_{n+1:N} \,|y_{1:n}\right)} = \frac{1}{S_{n+1}} \, \frac{p\left(y_{n+1:N} \,|x_n\right)}{p\left(y_{n+2:N} \,|y_{1:n+1}\right)}}.$$

- $n = N$:

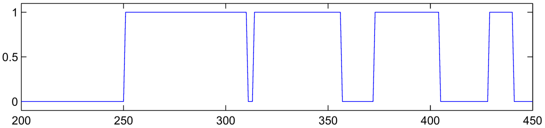
$$\beta_N(x_N) = 1.$$

- $1 \leq n < N$:

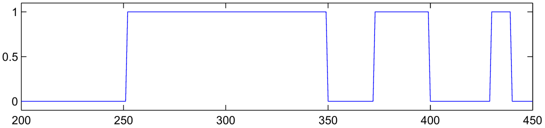
$$\beta_n(x_n) = \frac{1}{S_{n+1}} \sum_{x_{n+1}} \beta_{n+1}(x_{n+1}) \; p\left(z_{n+1} \,|x_n, y_{1:n}\right).$$

Résultat de restauration

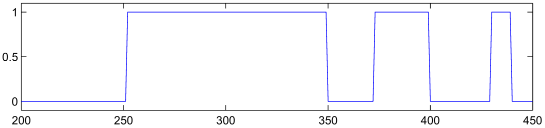
Restauration d'une CMC-BI (vrais paramètres)



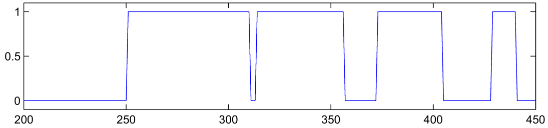
BI - $\tau = 8.4\%$



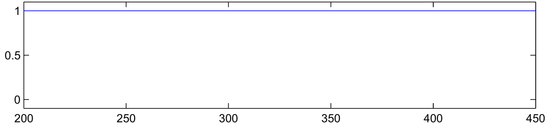
ML $(\alpha_1 = \alpha_2 = 10)$
$\tau = 8.4\%$



Restauration d'une CMC-ML (vrais paramètres)



BI - $\tau = $ **47.5%**



ML - $\tau = $ **6.1%**