



Modèles de Markov Couples et Triplets et quelques applications

F. Salzenstein

Université De Strasbourg
Laboratoire ICube, équipe IPP





- Dans un premier temps, définition des récursivités dans le cadre le plus général;
- Hypothèses markoviennes sur les données (observées, cachées), et markoviannité a posteriori qui en découlent;
- Hypothèses de d'indépendance/dépendance des données observées conditionnellement aux variables cachées;
- Hypothèse sur la nature du bruit: gaussien, non gaussien, à corrélation longue etc.;
- Hypothèses sur la nature des variables observées et/ou cachées: discrètes et/ou continues;
- Hypothèses sur l'aspect multi-dimensionnel;
- etc.

Hypothèses

- On définit les réalisations de la loi jointe de la variable aléatoire $Z = (X, Y)$ entre deux phénomènes, qui peuvent être un processus caché X et un processus observé Y ;
- Nécessité d'évaluer les probabilités *a posteriori* qui permettent l'estimation d'une réalisation de X à partir de celle de Y ;
- On se place dans le cadre de la segmentation bayésienne c'est à dire une stratégie de décision optimale vs une fonction de perte donnée;
- On peut supposer que les densités des observations conditionnelles aux données cachées, sont des modèles paramétriques dont il possible d'extraire les paramètres de forme Θ au moyen d'estimateurs du type EM, SEM, ICE;



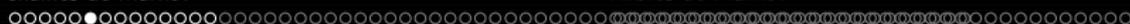
- Les estimateurs du MAP et du MPM sont donnés par:

$$\forall n \in \mathbb{N}, \hat{x}_n(y) = \sup_{x_n \in \Omega} p(x_n | y) \quad \hat{x}(y) = \sup_{x \in \Omega^N} p(x | y)$$

- L'estimation d'une réalisation de X nécessite le calcul des marginales a posteriori $p(x|y)$ et/ou $p(x_n|y)$ en chaque site;
- La loi *a posteriori* $p(x|y)$ peut être factorisée:

$$p(x|y) = p(x_N | x_{1:N-1}, y) p(x_{2:N-1} | y) = \dots = p(x_1 | y) \prod_{n=2}^N p(x_n | x_{1:n-1}, y) \quad (1)$$

- Le calcul nécessite la connaissance de la loi *a posteriori* initiale $p(x_1|y)$ et des lois de transition *a posteriori* $p(x_n|x_{1:n-1}, y)$;

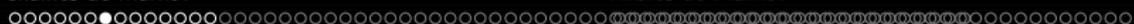


- Connaissant $p(x_1|y)$ et les lois de transition *a posteriori* le calcul de $p(x_n|y)$ peut s'effectuer récursivement:

$$\text{i) } p(x_{1:n}|y) = p(x_{1:n-1}|y)p(x_n|x_{1:n-1}, y)$$

$$\text{ii) } p(x_{n+1}|y) = \sum_{x_{1:n} \in \Omega^n} p(x_{1:n}|y)p(x_{n+1}|x_{1:n}, y)$$

- Les calculs deviennent prohibitifs avec la croissance de N . Les probabilités classiquement appelées “progressives” et “rétrogrades” permettent de contourner le problème.



Probabilités “progressives”

- On définit les quantités

$$\forall n \in \{1, \dots, N\}, \eta_n(x_{1:n}) = p(z_{1:n}) = p(x_{1:n}, y_{1:n});$$

- On pose de même: $\forall n \in \{1, \dots, N\}, \alpha_n(x_n) = p(x_n, y_{1:n});$

- Elles se calculent récursivement par une passe “avant” connaissant la loi $p(x_1)$ et les lois conditionnelles $p(z_{n+1}|z_{1:n})$:

$$\begin{aligned} \eta_1(x_1) &= p(x_1, y_1) = p(y_1|x_1)p(x_1) \\ \forall n \in \{1, \dots, N\} \quad \eta_{n+1}(x_{1:n+1}) &= p(x_{1:n+1}, y_{1:n+1}) \\ &= p(x_{n+1}, y_{n+1}|x_{1:n}, y_{1:n})\eta_n(x_{1:n}) \\ &= p(z_{n+1}|z_{1:n})\eta_n(x_{1:n}) \end{aligned}$$

- On déduit à chaque étape:

$$\alpha_{n+1}(x_{n+1}) = \sum_{x_{1:n} \in \Omega^n} \eta_{n+1}(x_{1:n+1}) \quad (2)$$



Probabilités "rétrogrades"

- On définit les quantités
 $\forall n \in \{1, \dots, N\}, \beta_n(x_{1:n}) = p(y_{n+1:N} | z_{1:n});$
- Elles se calculent récursivement en passe arrière:

$$\begin{aligned}
 \beta_n(x_{1:n}) &= \sum_{x_{n+1} \in \Omega} p(y_{n+2:N}, x_{n+1}, y_{n+1} | z_{1:n}) \\
 &= \sum_{x_{n+1} \in \Omega} p(y_{n+2:N} | z_{1:n+1}) p(z_{n+1} | z_{1:n}) \\
 &= \sum_{x_{n+1} \in \Omega} p(z_{n+1} | z_{1:n}) \beta_{n+1}(x_{1:n+1}) \quad (3)
 \end{aligned}$$

- En particulier:

$$\beta_{N-1}(x_{1:N-1}) = p(y_N | z_{1:N-1}) = \sum_{x_N \in \Omega} p(z_N | z_{1:N-1})$$

- Ceci implique: $\beta_N(x_{1:N}) = 1;$



Probabilités de transition *a posteriori*

- Les probabilités rétrogrades permettent d'exprimer $\forall n, p(x_n | x_{1:n-1}, y)$
- D'une part:

$$\begin{aligned}
 p(x_1 | y) &= \frac{p(x_1, y)}{\sum_{x_1 \in \Omega} p(x_1, y)} = \frac{p(y_{2:N} | x_1, y_1) p(x_1, y_1)}{\sum_{x_1 \in \Omega} p(y_{2:N} | x_1, y_1) p(x_1, y_1)} \\
 &= \frac{p(z_1) \beta(x_1)}{\sum_{x_1 \in \Omega} p(z_1) \beta(x_1)}
 \end{aligned}$$



Probabilités de transition *a posteriori* [Abbassi, 2012]

- Et $\forall n \in \{1, \dots, N - 1\}$:

$$\begin{aligned}
 p(x_{n+1} | x_{1:n}, y) &= p(x_{n+1} | x_{1:n}, y_{1:n}, y_{n+1:N}) \\
 &= \frac{p(x_{n+1}, y_{n+1:N} | x_{1:n}, y_{1:n})}{p(y_{n+1:N} | x_{1:n}, y_{1:n})} \\
 &= \frac{p(x_{n+1}, y_{n+1:N} | x_{1:n}, y_{1:n})}{\sum_{x_{n+1} \in \Omega} p(x_{n+1}, y_{n+1:N} | x_{1:n}, y_{1:n})} \\
 &= \frac{p(y_{n+2:N}, z_{n+1} | z_{1:n})}{\sum_{x_{n+1} \in \Omega} p(y_{n+2:N}, z_{n+1} | z_{1:n})} \\
 &= \frac{p(z_{n+1} | z_{1:n}) \beta_{n+1}(x_{1:n+1})}{\sum_{x_{n+1} \in \Omega} p(z_{n+1} | z_{1:n}) \beta_{n+1}(x_{1:n+1})}
 \end{aligned}$$

- Finalelement:

$$p(x_{n+1} | x_{1:n}, y) = p(z_{n+1} | z_{1:n}) \frac{\beta_{n+1}(x_{1:n+1})}{\beta_n(x_{1:n})}$$

Probabilités *a posteriori*

En utilisant l'expression (1), il vient:

$$p(x|y) = p(x_1|y) \prod_{n=2}^N p(x_n|x_{1:n-1}, y) = p(x_1|y) \prod_{n=2}^N p(z_n|z_{1:n-1}) \frac{\beta_n(x_{1:n})}{\beta_{n-1}(x_{1:n-1})}$$



Probabilités marginales *a posteriori* II

- De plus:

$$\begin{aligned}
 p(y) &= \sum_{x \in \Omega^N} p(x, y) \\
 &= \sum_{x_{1:n} \in \Omega^n} \sum_{x_{n+1:N} \in \Omega^{N-n}} p(x_{n+1:N} | x_{1:n}, y) \beta_n(x_{1:n}) \eta_n(x_{1:n}) \\
 &= \sum_{x_{1:n} \in \Omega^n} \beta_n(x_{1:n}) \eta_n(x_{1:n})
 \end{aligned}$$

- Ceci permet de calculer différemment $p(x|y)$:

$$p(x|y) = \frac{p(x, y)}{p(y)} = p(x_{n+1:N} | x_{1:n}, y) \frac{\beta_n(x_{1:n}) \eta_n(x_{1:n})}{\sum_{x_{1:n} \in \Omega^n} \beta_n(x_{1:n}) \eta_n(x_{1:n})}$$



Probabilités marginales *a posteriori* III

- Finalement:

$$\begin{aligned} \forall n \in \{1, \dots, N\}, p(x_n|y) &= \sum_{x_{1:n-1} \in \Omega^{n-1}} p(x_{1:n}|y) \\ &= \sum_{x_{1:n-1} \in \Omega^{n-1}} \sum_{x_{n+1:N} \in \Omega^{N-n}} p(x|y) \end{aligned}$$

- Soit, en utilisant (5):

$$\forall n \in \{1, \dots, N\}, p(x_n|y) = \sum_{x_{1:n-1} \in \Omega^{n-1}} \frac{\beta_n(x_{1:n}) \eta_n(x_{1:n})}{\sum_{x_{1:n} \in \Omega^n} \beta_n(x_{1:n}) \eta_n(x_{1:n})} \quad (6)$$

- Simplification de l'algorithme "forward" d'après (2):

$$\begin{aligned}
 \alpha_{n+1}(x_{n+1}) &= \sum_{x_{1:n} \in \Omega^n} \eta_{n+1}(x_{1:n+1}) \\
 &= \sum_{x_n \in \Omega} \sum_{x_{1:n-1} \in \Omega^{n-1}} p(z_{n+1} | z_{1:n}) \eta_n(x_{1:n}) \\
 &= \sum_{x_n \in \Omega} p(z_{n+1} | z_n) \alpha_n(x_n) \\
 &= p(y_{n+1} | x_{n+1}) \sum_{x_n \in \Omega} p(x_{n+1} | x_n) \alpha_n(x_n)
 \end{aligned}$$



- Simplification de la probabilité “backward”, $\forall n \in \mathcal{N}$:

$$\begin{aligned}\beta_n(\mathbf{x}_{1:n}) &= p(y_{n+1:N} | \mathbf{z}_{1:n}) \\ &= p(y_{n+1:N} | \mathbf{z}_{1:n-1}, \mathbf{x}_n, y_n) \\ &= p(y_{n+1:N} | \mathbf{x}_n) \text{ puisque } Y_{n+1:N} \perp\!\!\!\perp (Z_{1:n-1}, Y_n) \mid X_n \\ &= \beta_n(\mathbf{x}_n)\end{aligned}$$

- Simplification de l'algorithme “backward” d'après (3):

$$\begin{aligned}\beta_n(\mathbf{x}_n) &= \sum_{\mathbf{x}_{n+1} \in \Omega} p(z_{n+1} | z_n) \beta_{n+1}(\mathbf{x}_{n+1}) \\ &= \sum_{\mathbf{x}_{n+1} \in \Omega} p(y_{n+1} | \mathbf{x}_{n+1}) p(\mathbf{x}_{n+1} | \mathbf{x}_n) \beta_{n+1}(\mathbf{x}_{n+1})\end{aligned}$$



Marginale *a posteriori*

- D'après (6):

$$\begin{aligned} \forall n \in \{1, \dots, N\}, p(x_n | y) &= \frac{\sum_{x_{1:n-1} \in \Omega^{n-1}} \beta_n(x_n) \eta_n(x_{1:n})}{\sum_{x_n \in \Omega} \sum_{x_{1:n-1} \in \Omega^{n-1}} \beta_n(x_n) \eta_n(x_{1:n})} \\ &= \frac{\beta_n(x_n) \alpha_n(x_n)}{\sum_{x_n \in \Omega} \beta_n(x_n) \alpha_n(x_n)} \end{aligned}$$

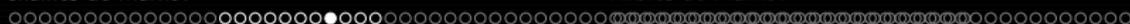
Markoviannité de X *a posteriori*

- D'après (4):

$$p(x_{n+1}|x_{1:n}, y) = p(z_{n+1}|z_n) \frac{\beta_{n+1}(x_{n+1})}{\beta_n(x_n)}$$

- Les termes $x_{1:n-1}$ ne jouent aucun rôle, donc:

$$p(x_{n+1}|x_{1:n}, y) = p(x_{n+1}|x_n, y)$$



Markoviannité de X *a posteriori*

- Ceci permet la simulation d'une chaîne a posteriori pour segmenter et/ou estimer les hyper-paramètres $\theta^{(q+1)}$ des modèles récursivement selon les hyper-paramètres courant $\theta^{(q)}$;
- Loi de transition *a posteriori*:

$$p(x_{n+1}|x_n, y, \theta^{(q)}) = p(y_{n+1}|x_{n+1})p(x_{n+1}|x_n) \frac{\beta_{n+1}(x_{n+1})}{\beta_n(x_n)}$$

- Loi initiale *a posteriori*:

$$p(x_n|y, \theta^{(q)}) \propto \alpha_n(x_n)\beta_n(x_n)$$

- Par la formule de Bayes, loi conjointe *a posteriori*:

$$p(x_n, x_{n+1}|y, \theta^{(q)}) \propto \alpha_n(x_n)\beta_{n+1}(x_{n+1})p(y_{n+1}|x_{n+1})p(x_{n+1}|x_n)$$



Algorithme de Devijver [Devijver *et al.*, 1985] I

- Assure la stabilité numérique de l'algorithme "forward-backward";
- Probabilités "forward" :

$$\begin{aligned}
 \tilde{\alpha}_{n+1}(x_{n+1}) &= p(x_{n+1}|y_{1:n+1}) = \frac{\alpha_{n+1}(x_{n+1})}{\sum_{x_{n+1} \in \Omega} \alpha_{n+1}(x_{n+1})} \\
 &= \frac{p(y_{n+1}|x_{n+1}) \sum_{x_n \in \Omega} p(x_{n+1}|x_n) p(x_n|y_{1:n}) p(y_{1:n})}{\sum_{x_{n+1} \in \Omega} p(y_{n+1}|x_{n+1}) \sum_{x_n \in \Omega} p(x_{n+1}|x_n) p(x_n|y_{1:n}) p(y_{1:n})} \\
 &= \frac{p(y_{n+1}|x_{n+1}) \sum_{x_n \in \Omega} p(x_{n+1}|x_n) \tilde{\alpha}_n(x_n)}{\sum_{x_{n+1} \in \Omega} p(y_{n+1}|x_{n+1}) \sum_{x_n \in \Omega} p(x_{n+1}|x_n) \tilde{\alpha}_n(x_n)}
 \end{aligned}$$

Algorithme de Devijver [Devijver *et al.*, 1985] II

- Probabilité “backward” :

$$\tilde{\beta}_n(x_n) = \frac{p(y_{n+1:N}|x_n)}{p(y_{n+1:N}|y_{1:n})}$$

- Probabilité *a posteriori*:

$$\forall n \in \{1, \dots, N\}, p(x_n|y) \propto \tilde{\beta}_n(x_n)\tilde{\alpha}_n(x_n)$$



Processus observé à mémoire longue [Abbassi, 2012]

- Hypothèse markovienne sur X ;
- Les variables Y_n vérifient la propriété:

$$\forall n \in \mathcal{N}, p(y_n | x_{1:n}, y_{1:n-1}) = p(y_n | x_n, y_{1:n-1})$$

- Lorsque les lois $p(y_n | x_n, y_{1:n-1})$ sont calculables, les inférences sont possibles:

$$\begin{aligned} p(z_{n+1} | z_{1:n}) &= p(x_{n+1}, y_{n+1} | x_{1:n}, y_{1:n}) \\ &= p(y_{n+1} | x_{1:n}, y_{1:n}, x_{n+1}) p(x_{n+1} | x_{1:n}, y_{1:n}) \\ &= p(y_{n+1} | x_{n+1}, y_{1:n}) p(x_{n+1} | x_n) \end{aligned}$$

Markoviannité *a posteriori* de X (resp. Y)

- Proposition [Pieczynski, 2003]:

- ▶ Soit $Z_n = (X_n, Y_n)$ une CM couple, alors:
- ▶ X conditionnellement à Y est une chaîne de Markov;
- ▶ Y conditionnellement à X est une chaîne de Markov;

- Démonstration [Pieczynski, 2003] :

$$\begin{aligned}
 p(x|y) &= \frac{p(z)}{p(y)} = \frac{p(z_1)}{p(y)} \prod_{n=2}^N p(z_n|z_{n-1}) \\
 &= \frac{p(x_1, y_1)p(x_2, y_2|x_1, y_1)\dots p(x_n, y_n|x_{n-1}, y_{n-1})}{p(y)}
 \end{aligned}$$

Lorsque y est constante dans l'égalité précédente, on reconnaît la factorisation markovienne de $p(x|y)$.

Conditions de markoviannité *a priori* de X

- Dans le cas de chaînes stationnaires réversibles, sous certaines conditions une CM couple est une CMC;
- Proposition [Pieczynski, 2003]:

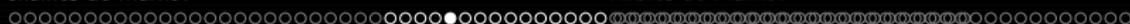
Soit $Z_n = (X_n, Y_n)$ une CM couple, qui vérifie:

- $p(z_n, z_{n+1})$ ne dépend pas de n , $\forall n \in \mathcal{N}$
- Z est réversible *i.e.* $p(z_n, z_{n+1}) = p(z_{n+1}, z_n)$

Alors les trois conditions suivantes sont équivalentes:

- X est une chaîne de Markov;
- $\forall n, p(y_n | x_n | x_{n-1}) = p(y_n | x_n)$
- $\forall n, p(y_n | x) = p(y_n | x_n)$

- Pour la démonstration voir [Pieczynski, 2003], [Ben Mabrouk, 2011].



Loi de transition de Z

- Le calcul de l'inférence bayésienne nécessite la connaissance des probabilités de transition:

$$p(z_{n+1}|z_n) = \frac{p(z_{n+1}, z_n)}{p(z_n)} = \frac{p(x_n, x_{n+1})p(y_n, y_{n+1}|x_n, x_{n+1})}{\sum_{x_{n+1} \in \Omega} p(x_n, x_{n+1})p(y_n|x_n, x_{n+1})} \quad (7)$$

- Ou en fonction des probabilités de transition de X le cas échéant:

$$\begin{aligned} p(z_{n+1}|z_n) &= \frac{p(z_{n+1}, z_n)}{p(z_n)} = \frac{p(x_n, x_{n+1})p(y_n, y_{n+1}|x_n, x_{n+1})}{p(x_n, y_n)} \\ &= \frac{p(x_{n+1}|x_n)p(y_n, y_{n+1}|x_n, x_{n+1})}{p(y_n|x_n)} \end{aligned} \quad (8)$$



CMCouples particuliers II

- Dans le cas stationnaire, la connaissance de la densité $p(z_1, z_2)$ permet les calculs;
- Dans le cas CMC-BI, d'après (8):

$$p(z_{n+1}|z_n) = p(y_{n+1}|x_{n+1})p(x_{n+1}|x_n)$$

- Dans le cas d'une CMC les y_n sont dépendants, nous avons la condition $p(y_n|x_n, x_{n-1}) = p(y_n|x_n)$, et d'après (8):

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

CMCouples particuliers III

- Dans le cas d'une CM Couple BI les y_n sont indépendants conditionnellement à x , nous avons la condition:

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

- Les transitions sont données par:

$$p(z_{n+1} | z_n) = p(x_{n+1} | z_n)p(y_{n+1} | x_n, x_{n+1})$$



CMCouples avec observations cond. dépendantes

- D'après (7) le calcul de la probabilité de transition $p(z_{n+1}|z_n)$ nécessite la connaissance des probabilités $p(x_n, x_{n+1})$, $p(y_n, y_{n+1}|x_n, x_{n+1})$ et $p(y_n|x_n, x_{n+1})$;
- La probabilité $p(x_n, x_{n+1})$ peut être estimée à partir de la simulation d'une réalisation $X = x$ de la chaîne a posteriori sachant $Y = y$;
- La probabilité $p(y_n|x_n, x_{n+1})$ peut se calculer en posant:

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

- Des modèles de densité appropriés permettent de calculer $p(y_{n+1}|y_n, x_n, x_{n+1})$ et $p(y_n, y_{n+1}|x_n, x_{n+1})$.

CMCouples avec données corrélées gaussiennes

- Densité multivariée $p(y_n, y_{n+1} | x_n, x_{n+1})$:

$$p(y_n, y_{n+1} | (\omega_i, \omega_j)) = \frac{1}{2\pi \det(\Gamma_{ij})} \exp -\frac{1}{2} (\mathbf{y}_{n,n+1} - \boldsymbol{\mu}_{ij})^T \Gamma_{ij}^{-1} (\mathbf{y}_{n,n+1} - \boldsymbol{\mu}_{ij})$$

- En posant $\mathbf{y}_{n,n+1} = (y_n, y_{n+1})^T$. Avec les matrices moyennes et variances-covariances:

$$\boldsymbol{\mu}_{i,j} = \begin{pmatrix} \mu_{i,j}^{(1)} \\ \mu_{i,j}^{(2)} \end{pmatrix}; \quad \Gamma_{i,j} = \begin{pmatrix} \Gamma_{i,j}^{(1,1)} & \Gamma_{i,j}^{(1,2)} \\ \Gamma_{i,j}^{(2,1)} & \Gamma_{i,j}^{(2,2)} \end{pmatrix}$$

- Par propriétés des gaussiennes:

$p(y_{n+1} | y_n, (x_n, x_{n+1})) \sim \mathcal{N}(m_{ij}, \sigma_{ij})$ où les moyennes et variances $(m_{ij}; \sigma_{ij})$ sont données par:

$$m_{i,j} = \mu_{i,j}^{(1)} + \frac{\Gamma_{i,j}^{(1,2)}}{\Gamma_{i,j}^{(2,2)}} \cdot (y_n - \mu_{i,j}^{(2)}); \quad \sigma_{i,j} = \Gamma_{i,j}^{(1,1)} - \frac{\Gamma_{i,j}^{(1,2)} \cdot \Gamma_{i,j}^{(2,1)}}{\Gamma_{i,j}^{(2,2)}}$$

Inférence bayésienne dans les CM Couples I

- Simplification de l'algorithme "forward" d'après (2):

$$\begin{aligned}
 \alpha_{n+1}(x_{n+1}) &= \sum_{x_{1:n} \in \Omega^n} \eta_{n+1}(x_{1:n+1}) \\
 &= \sum_{x_n \in \Omega} \sum_{x_{1:n-1} \in \Omega^{n-1}} p(z_{n+1} | z_{1:n}) \eta_n(x_{1:n}) \\
 &= \sum_{x_n \in \Omega} p(z_{n+1} | z_n) \alpha_n(x_n)
 \end{aligned}$$

- Simplification de la probabilité "backward":

$$\forall n \in \mathcal{N}, \beta_n(x_{1:n}) = p(y_{n+1:N} | z_{1:n}) = p(y_{n+1:N} | x_n) = \beta_n(x_n)$$

- Simplification de l'algorithme "backward" d'après (3):

$$\beta_n(x_n) = \sum_{x_{n+1} \in \Omega} p(z_{n+1} | z_n) \beta_{n+1}(x_{n+1})$$

Inférence bayésienne dans les CM Couples II

- Segmentation MPM par la relation, d'après (6) qui s'applique facilement au cas couple:

$$p(x_n|y) \propto \alpha_n(x_n)\beta_n(x_n)$$

- Simulation d'une chaîne de Markov $X = \hat{x}$ a posteriori d'après la relation (4):

$$p(x_{n+1}|x_n, y) = p(z_{n+1}|z_n) \frac{\beta_{n+1}(x_{n+1})}{\beta_n(x_n)}$$

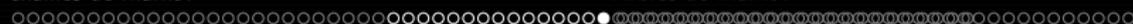
- Les réalisations $X = \hat{x}$ permettent la mise en oeuvre d'algorithmes d'estimation des hyper-paramètres (EM, SEM, ICE).

Estimation des hyper-paramètres par SEM I

- l'algorithme final de segmentation nécessite la connaissance de $\theta = (\theta_X, \theta_Y)$;
- Le paramètre *a priori* θ_X définissent les probabilités (ev. densités) conjointes $p(x_n, x_{n+1})$ de $X = x$.
- Les paramètres $\theta_Y = (\mu_{ij}, \Gamma_{ij})_{i,j \in \{1, \dots, K\}}^2$ définissent les vraisemblances (Y_n, Y_{n+1}) conditionnellement à (X_n, X_{n+1}) i.e, ceux de la densité bivariee.
- En supposant la stationnarité et l'homogénéité du processus de Markov *a posteriori*, $X|Y$ permet l'emploi du SEM: observant une réalisation de X selon $p(X|Y = y)$, nous définissons l'estimateur $\hat{\theta} = (\hat{\theta}_X, \hat{\theta}_Y) = \hat{\theta}(X, Y)$ des données complètes (X, Y) .
- Dans [Lanchantin, 2006] développement complet des algorithmes EM, SEM, ICE pour les CM Couples.

Estimation des hyper-paramètres par SEM II

- Initialisation $\theta^{(0)} = (\theta_X^{(0)}, \theta_Y^{(0)})$.
- Étape courante:
 - 1 Simulation d'une réalisation de $x^{(q)}$ selon $p_{\theta^{(q)}}(X|Y=y)$
 - 2 Estimation des nouveaux paramètres $\theta^{(q+1)} = \hat{\theta}(x^{(q)}, y)$ d'après $(x^{(q)}, y)$:
- Si $\theta^{(q+1)}$ est stable, stopper la procédure.



Exemple: cas gaussien [Le Cam *et al.*, 2008]

- Soit A^i l'ensemble des singletons $\{x_n\}_{1 \leq n \leq N}$ de $X^{(q)}$ où $x_n = \omega_i, 1 \leq i \leq K$;
- Soit $A^{i,j}$ l'ensemble des voisinages $\{x_n, x_{n+1}\}_{1 \leq n \leq N-1}$ de $X^{(q)}$ où $(x_n, x_{n+1}) = (\omega_i, \omega_j)$;
- Paramètres de vraisemblance au moyen des fonctions indicatrices $\mathbb{1}_{A^{i,j}}$:

$$\mu_{ij}^{(q)} = \frac{\sum_{n=1}^{N-1} \mathbb{1}_{A^{i,j}} \mathbf{y}_{n,n+1}}{\text{card}(A^{i,j})} \quad (9)$$

$$\Gamma_{ij}^{(q)} = \frac{\sum_{n=1}^{N-1} \mathbb{1}_{A^{i,j}} (\mathbf{y}_{n,n+1} - \mu_{ij}^{(q)}) (\mathbf{y}_{n,n+1} - \mu_{ij}^{(q)})^T}{\text{card}(A^{i,j})} \quad (10)$$

- Loi conjointe de X à partir d'une réalisation $x^{(q)}$ de X selon $p_{\theta^{(q)}}(X|Y=y)$.

$$p(x_n = \omega_i, x_{n+1} = \omega_j | \theta^{(q+1)}) = \frac{\text{card}(A^{i,j})}{N}$$

- Les chaînes couples partiellement de Markov [Pieczysnki, 2004] consistent à poser l'hypothèse d'indépendance:

$$Z_{n+1:N} \perp\!\!\!\perp X_{1:n-1} \mid (Z_n, Y_{1:n-1})$$

- Dans ce modèle ni Z , ni X ne sont des CM. Toutefois la loi a posteriori de $X \mid Y = y$ est celle d'une CM.
- Loi de transition $p(z_{n+1} \mid z_{1:n})$:

$$\begin{aligned}
 p(z_{n+1} \mid z_{1:n}) &= p(z_{n+1} \mid z_n, x_{1:n-1}, y_{1:n-1}) \\
 &= p(z_{n+1} \mid z_n, y_{1:n-1}) \\
 &= p(x_{n+1} \mid z_n, y_{1:n-1}) p(y_{n+1} \mid x_{n+1}, z_n, y_{1:n-1})
 \end{aligned}$$

Markoviannité de X *a posteriori*

- Loi “rétrograde”:

$$\begin{aligned}
 \beta_n(x_{1:n}) &= p(y_{n+1:N} | z_{1:n}) \\
 &= p(y_{n+1:N} | x_{1:n-1}, x_n, y_{1:n-1}, y_n) \\
 &= p(y_{n+1:N} | z_n, y_{1:n-1},) \text{ puisque } Y_{n+1:N} \perp\!\!\!\perp X_{1:n-1} | (Z_n, Y_{1:n-1}) \\
 &= \beta_n(x_n, y_{1:n})
 \end{aligned}$$

- D'après (4) et ce qui précède:

$$\begin{aligned}
 p(x_{n+1} | x_{1:n}, y) &= p(z_{n+1} | z_n, y_{1:n-1}) \frac{\beta_{n+1}(x_{n+1}, y_{1:n+1})}{\beta_n(x_n, y_{1:n})} \\
 &= p(x_{n+1} | x_n, y)
 \end{aligned}$$

- Ces probabilités de transition sont calculables récursivement si les probabilités $p(z_{n+1} | z_n, y_{1:n-1})$ sont calculables;

- Décomposition de $p(z_{n+1}|z_n, y_{1:n-1})$:

$$\begin{aligned}
 p(z_{n+1}|z_n, y_{1:n-1}) &= \frac{p(z_n, z_{n+1}, y_{1:n-1})}{p(z_n, y_{1:n-1})} \\
 &= \frac{p(x_n, x_{n+1})p(y_{1:n+1}|x_n, x_{n+1})}{\sum_{x_{n+1} \in \Omega} p(x_n, x_{n+1})p(y_{1:n}|x_n, x_{n+1})}
 \end{aligned}$$

- Dans certains cas (bruits à corrélation longue) les probabilités $p(y_{1:n+1}|x_n, x_{n+1})$ peuvent être calculées récursivement [Lanchantin, 2006] sous hypothèses gaussiennes ou non [Pieczynski, 2005].

- En posant $V = (X, U)$, le processus T est donc un processus couple $T = (V, Y)$ (resp. $T = (V, X)$);
- En posant $Z = (X, Y)$ on considère $T = (Z, U)$;
- Le processus $T = (X, U, Y)$ peut être markovien *i.e.*:

$$p(t) = p(t_1) \prod_{n=2}^N p(t_n | t_{n-1})$$

- Il est possible d'appliquer les algorithmes précédents en traitant par exemple $T = (V, Y)$ comme un processus couple.

- Calcul des marginales *a posteriori* $p(v_n|y) = p(x_n, u_n|y)$ par l'algorithme "forward-backward" :

$$\alpha_1(v_1) = p(v_1) \quad \alpha_{n+1}(v_{n+1}) = \sum_{v_n \in \Omega \times \Lambda} \alpha_n(v_n) p(t_{n+1}|t_n) \quad (11)$$

$$\beta_N(v_N) = 1 \quad \beta_n(v_n) = \sum_{v_{n+1} \in \Omega \times \Lambda} \beta_{n+1}(v_{n+1}) p(t_{n+1}|t_n) \quad (12)$$

- Les marginales *a posteriori* du processus caché X sont déduites par:

$$p(x_n|y) = \sum_{u_n \in \Lambda} p(x_n, u_n|y)$$

Quelques généralisations des modèles précédents

- Lorsque $V = (X, U)$ (resp. $Z = (X, Y)$) est une chaîne de Markov, on a affaire à une chaîne de Markov couple cachée dont la loi de transition est donnée par:

$$p(t_{n+1}|t_n) = p(v_{n+1}|v_n)p(y_n|v_n)$$

- Condition nécessaire et suffisante pour que V soit une CM:

$$\forall n \ p(y_n|x_n, u_n, x_{n+1}, u_{n+1}) = p(y_n|x_n, u_n)$$

- Lorsque les hypothèses d'indépendance suivantes sont vérifiées, on a affaire à une chaîne triplet partiellement de Markov:

$$T_{n+1:N} \perp\!\!\!\perp V_{1:n-1} | (T_n, Y_{1:n-1})$$



Chaîne de Markov M-stationnaire

- Une chaîne de Markov stationnaire est une chaîne de Markov pour laquelle les lois de transition ne dépendent pas de n ;
- L'idée du modèle [Lanchantin, 2004] est de modéliser la non stationnarité du processus caché en utilisant un processus auxiliaire U ;
- Le processus auxiliaire U modélise les changements des matrices de transition;
- Le processus $Z = (X, Y)$ est une chaîne de Markov couple M-non stationnaire, s'il existe $U \in \Lambda^N$, $|\Lambda| = M$ tel que la chaîne $T = (X, U, Y)$ est une chaîne de Markov stationnaire;
- En général, Z n'est ni markovienne, ni stationnaire;
- Le processus $(X, Y) | U = u$ est markovien *a posteriori* et en général non stationnaire.



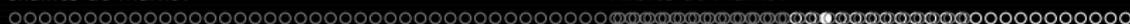
Sommaire

- 1 chaînes de Markov
 - Algorithme d'inférence général
 - Chaînes de Markov Cachées à bruit indépendant (CMC-BI)
 - Chaîne de Markov couples (CMCo)
 - Chaînes couples partiellement de Markov (CCPM)
 - Chaînes de Markov triplets (CMT)
 - Exemples
- 2 Arbres de Markov
 - Algorithme d'inférence général
 - Arbres de Markov caché à bruit indépendant (AMC-BI)
 - Arbre de Markov Couple (AMCo)
 - Arbre Couple partiellement de Markov (ACPM)
 - Exemples
- 3 Conclusion
 - Conclusion



Chaînes de Markov floues non stationnaires [Salzenstein et al., 2007]

- L'objectif consiste à estimer la réalisation cachée $x = (x_s)_{1 \leq s \leq N}$, étant données D observations $\mathbf{Y} = \mathbf{y} = \{\mathbf{y}_s \in \mathbb{R}^D\}$;
- En chaque site: $x_s = (\varepsilon_1(s), \varepsilon_2(s), \dots, \varepsilon_K(s))$;
- Chaque composant $\varepsilon_i(s)$ représente la contribution de la classe ω_i dans un ensemble discret $\Omega = \{\omega_1, \dots, \omega_K\}$ de K classes dites "dures";
- Condition de normalisation: $\varepsilon_1(s) + \varepsilon_2(s) + \dots + \varepsilon_K(s) = 1$;
- Dans le contexte de deux classes "dures": $\Omega = \{0, 1\}$ soit $x_s \in [0, 1]$;



Cas stationnaire: expression des densités

- La distribution en chaque X_s est donnée par une densité h_s selon une mesure ν incluant des composantes discrètes (Dirac δ_0, δ_1 sur $\{0, 1\}$) et une composante continue (mesure de Lebesgue μ on $]0, 1[$):

$$\nu = \delta_0 + \delta_1 + \mu \quad (13)$$

- Les composantes discrètes de ν sont associées aux classes “dures”, tandis que la/les composante(s) continue(s) μ est associée à la caractéristique floue;
- Lorsque X est une chaîne de Markov floues (FMC) avec les hypothèses classiques d'indépendances conditionnelles de \mathbf{Y} selon X , la distribution jointe $p(x, \mathbf{y})$ selon la mesure $\nu^N \otimes \mu^N$ est:

$$p(x, \mathbf{y}) = p(x_1)p(x_2|x_1)\dots p(x_N|x_{N-1})p(\mathbf{y}_1|x_1)\dots p(\mathbf{y}_N|x_N) \quad (14)$$

- La V.A. X a posteriori conditionnellement à \mathbf{Y} est une chaîne de Markov;
- Généralement la distribution $p(x, \mathbf{y})$ dépend des paramètres $\theta = (\theta_X, \theta_Y)$;
- Les paramètres *a priori* θ_X définissent la loi de la chaîne de Markov et θ_Y définit la vraisemblance des observations sachant X .
- Lorsque $X = (x_s)_{1 \leq s \leq N}$ est une FMC stationnaire *i.e.* $X_s \in [0, 1]$, la distribution *a priori* $\pi(x)$ de la V.A. X , est donnée par la loi initiale $p(x_1)$ et les densités de transition $p(x_s | x_{s-1})_{1 \leq s \leq N}$:

$$p(x_1, x_2, \dots, x_N) = p(x_1)p(x_2|x_1)\dots p(x_N|x_{N-1}) \quad (15)$$

- Lorsque la chaîne est stationnaire, la loi a priori est définie par la densité jointe $p(x_s, x_{s+1})$ étant donnés $(x_s, x_{s+1}) \in [0, 1]^2$.
- En fonction de la mesure $\nu \otimes \nu$, la condition de normalisation s'exprime:

$$\int_0^1 \int_0^1 p(u, v) d(\nu \otimes \nu)(u, v) = 1 \quad (16)$$

- Localement, le modèle de densité proposé est le suivant:

$$p(\varepsilon_1, \varepsilon_2) = \begin{cases} a(1 - |\varepsilon_1 - \varepsilon_2|)^r + b; & r \in \mathbb{R} \quad \text{si } \varepsilon_1 \text{ ou } \varepsilon_2 \in]0, 1[\\ \pi_{ij} & \text{si } (i, j) \in \{0, 1\}^2 \end{cases}$$

Modèle CMC-MS-BI appliqué au cas flu

- Au processus initial X - non nécessairement markovien - on associe un processus $U \in \Lambda^N$, avec $\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_K\}$;
- Le processus couple $Z = (X, U)$ est supposé être une CMC stationnaire:

$$p(z_s = (x_s, u_s) | z_{s-1}, z_{s-2}, \dots, z_1) = p(z_s | z_{s-1}) \quad (17)$$

- La chaîne Z est totalement définie par:

$$p(z_s, z_{s+1}) = p(x_s, x_{s+1}, u_s, u_{s+1})$$

- En posant $I_s \subset [0, 1]$ et $I_{s+1} \subset [0, 1]$, selon la mesure $(\nu + \sum_{n=1}^K \delta_{\lambda_n}) \otimes (\nu + \sum_{n=1}^K \delta_{\lambda_n})$, la probabilité initiale est donnée par:

$$\begin{aligned}
 & P(X_s \in I_s, X_{s+1} \in I_{s+1}, U_s = \lambda_s, U_{s+1} = \lambda_{s+1}) \\
 &= \int_{I_s} \int_{I_{s+1}} p(\epsilon, \eta, \lambda_s, \lambda_{s+1}) d(\nu \otimes \nu)(\epsilon, \eta)
 \end{aligned} \quad (18)$$

Paramètres *a priori* I

- Condition de normalisation (19):

$$\sum_{\lambda_i} \sum_{\lambda_j} \int_0^1 \int_0^1 p(\epsilon, \eta, \lambda_i, \lambda_j) d(\nu \otimes \nu)(\epsilon, \eta) = 1 \quad (19)$$

- En particulier, cette condition s'exprime:

$$\sum_{\lambda_i} \sum_{\lambda_j} P_{ij} = 1 \quad (20)$$

- Avec

$$\int_0^1 \int_0^1 p(\epsilon, \eta, \lambda_i, \lambda_j) d(\nu \otimes \nu)(\epsilon, \eta) = P[\lambda_i, \lambda_j] = P_{ij}$$



Paramètres *a priori* II

- On construit une densité paramétrée de la forme suivante:

$$p(\epsilon_1, \epsilon_2, \lambda_i, \lambda_j) = a_{ij}\phi(\epsilon_1, \epsilon_2) + b_{ij} \quad (22)$$

- Avec les conditions aux limites (23):

$$a_{ij} \cdot \phi(0, 1) + b_{ij} = a_{ij} \cdot \phi(1, 0) + b_{ij} = b_{ij} = \pi_{01}^{ij} = \pi_{10}^{ij} \quad (23)$$

- Au total, il y a $6K^2$ paramètres de la loi *a priori*.

Vraisemblance des données multi-spectrales I

- Nous observons D réalisations $(y^{(1)}, y^{(2)}, \dots, y^{(D)})$ du vecteur $\mathbf{Y} = (Y^{(1)}, Y^{(2)}, \dots, Y^{(D)})$;
- Pour chaque $\mathbf{Y}^{(i)}$, les V.A. $\mathbf{Y}^{(i)} = \{Y_1^{(i)}, Y_2^{(i)}, \dots, Y_N^{(i)}\}$ sont indépendantes conditionnellement à Z ;
- Hypothèses suivantes [Lachantin, 2004]:

$$p(\mathbf{Y}^{(i)}|Z) = \prod_{s=1}^N p(\mathbf{Y}^{(i)}_s|Z) \quad (24)$$

$$p(\mathbf{Y}^{(i)}_s|Z) = p(\mathbf{Y}^{(i)}_s|Z_s) \quad (25)$$

$$p(\mathbf{Y}^{(i)}_s|Z_s) = p(\mathbf{Y}^{(i)}_s|X_s) \quad (26)$$

- Ici, le paramètre θ_X varie localement, tandis que θ_Y reste global.

Vraisemblance des données multi-spectrales

- La distribution $f_{X_s}(\mathbf{y}_s)$ de \mathbf{y}_s sachant $X_s = \varepsilon_s \in [0, 1]$ est une gaussienne multi-variée:

$$f_{X_s}(\mathbf{y}_s) = \frac{1}{2\pi^{D/2}(\det \Gamma_{\varepsilon_s})^{1/2}} \exp\left(\frac{-1}{2}(\mathbf{y}_s - \mu_{\varepsilon_s})^t \Gamma_{\varepsilon_s}^{-1}(\mathbf{y}_s - \mu_{\varepsilon_s})\right) \quad (27)$$

- Soient (μ_0, μ_1) et (Γ_0, Γ_1) les moyennes et matrices variance-covariance relativement aux classes "dures" "0" and "1";
- En chaque site $X_s = \varepsilon_s$, le vecteur moyenne et la matrice de covariance μ_{ε_s} et Γ_{ε_s} s'expriment:

$$\mu_{\varepsilon_s} = (1 - \varepsilon_s) \cdot \mu_0 + \varepsilon_s \mu_1 \quad (28)$$

$$\Gamma_{\varepsilon_s} = (1 - \varepsilon_s)^2 \cdot \Gamma_0 + \varepsilon_s^2 \Gamma_1 \quad (29)$$



Segmentation MPM

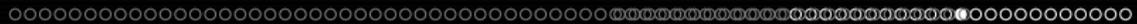
- Il est possible d'estimer X et U , en marginalisant la distribution $p(Z_s|y)$:

$$p(x_s|y) = \sum_{\Lambda} p(z_s|y) ; p(u_s|y) = \int_0^1 p(z_s|y) d\nu$$

- Procédure "forward-backward" étendue au contexte fluu:

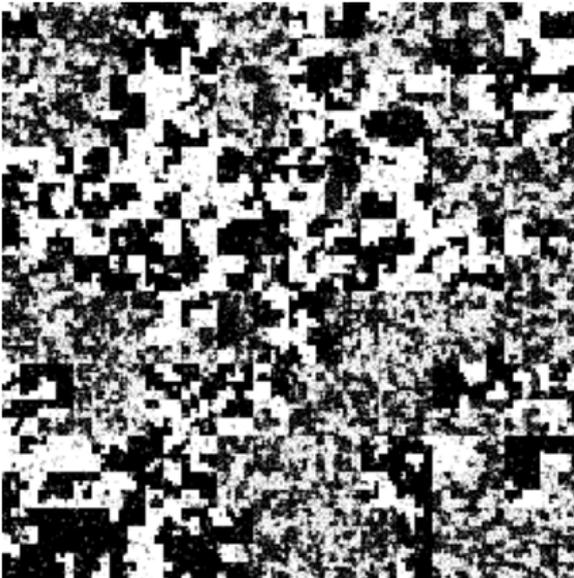
$$\alpha_s(z_s) \propto p(\mathbf{y}_s|z_s) \sum_{\lambda \in \Lambda} \int_0^1 \alpha_{s-1}(\varepsilon, \lambda) p(z_s|\varepsilon, \lambda) d\nu(\varepsilon)$$

$$\beta_s(z_s) \propto \sum_{\lambda \in \Lambda} \int_0^1 \beta_{s+1}(\varepsilon, \lambda) p(\mathbf{y}_{s+1}|\varepsilon, \lambda) p(\varepsilon, \lambda|z_s) d\nu(\varepsilon)$$



Simulations

Chaîne non stationnaire



(a)

Etats auxiliaires



(b)

Figure: (a): Chaîne non stationnaire X ; (b): États U

Exemples

Image réelle (Oakland)



(a)



(b)

Figure: (a): Observation; (b) Champ X



Lissage exact dans un modèle à sauts

- On considère $X_1^N = (X_1, X_2, \dots, X_N)$, $R_1^N = (R_1, R_2, \dots, R_N)$ et $Y_1^N = (Y_1, Y_2, \dots, Y_N)$ trois processus tels que X est à valeurs dans \mathbb{R}^q , R à valeurs dans un ensemble fini Λ^K , et Y à valeurs dans \mathbb{R}^m ;
- Le processus R modélise les sauts de la distribution de (X, Y) ;
- Le lissage consiste à calculer $\forall n$, l'espérance conditionnelle $E[X_n | Y_1^N = y_1^N]$ noté $E[X_n | y_1^N]$;
- Le modèle classique gaussien consiste à poser:

$R = (R_n)_{n \in \mathcal{N}}$ est une chaîne de Markov;

$$X_n = F_n(R_n) X_{n-1} + G_n(R_n) W_n$$

$$Y_n = H_n(R_n) X_n + J_n(R_n) V_n$$



- Généralement W_n et V_n sont indépendantes et gaussiennes;
- Les matrices $F_n(R_n)$, $G_n(R_n)$, $H_n(R_n)$ et $J_n(R_n)$ dépendent de R_n ;
- Dans ces conditions la loi conditionnelle de $(X, Y|R = r)$ est gaussienne;
- On note la propriété:

$$p(x_{n+1}, r_{n+1} | x_n, r_n, y_n) = p(x_{n+1}, r_{n+1} | x_n, r_n)$$

- Lorsque R est caché, le calcul de $E[X_n | y_1^N]$ devient prohibitif;
- Soit dans ces conditions (processus R caché) le modèle suivant:

(R_1^N, Y_1^N) est une chaîne de Markov;

$$X_{n+1} = F_{n+1}(R_{n+1}, Y_{n+1}) X_n + G_{n+1}(R_{n+1}, Y_{n+1}) W_{n+1}$$

- Dans ce cas, le processus $T = (X, R, Y)$ est markovien avec:

$$p(y_{n+1}, r_{n+1} | x_n, r_n, y_n) = p(y_{n+1}, r_{n+1} | y_n, r_n)$$



- Dans le premier modèle (R, X) est markovien mais pas (R, Y) : situation inverse dans le second modèle;
- On montre que la markoviannité de (R, Y) permet de calculer $E[X_{n+1}|r_{n+1}, y_1^N]$ et $p(r_{n+1}|y_1^N)$ à partir $E[X_n|r_n, y_1^N]$ et $p(r_n|y_1^N)$;
- Considérons l'espérance de (30) conditionnellement à $(R_{n+1}, Y_1^N) = (r_{n+1}, y_1^N)$:

$$\begin{aligned}
 E[X_{n+1}|r_{n+1}, y_1^N] &= F_{n+1}(r_{n+1}, y_{n+1}) E[X_n|r_{n+1}, y_1^N] \\
 &= \sum_{r_n} F_{n+1}(r_{n+1}, y_{n+1}) E[X_n, r_n|r_{n+1}, y_1^N] \\
 &= \sum_{r_n} F_{n+1}(r_{n+1}, y_{n+1}) E[X_n|r_n, r_{n+1}, y_1^N] p(r_n|r_{n+1}, y_1^N) \\
 &= \sum_{r_n} F_{n+1}(r_{n+1}, y_{n+1}) E[X_n|r_n, y_1^N] p(r_n|r_{n+1}, y_1^N)
 \end{aligned}$$

- Les lois marginales et les transitions *a posteriori* de $r_n | Y_1^N = y_1^N$ peuvent être calculées par les algorithmes classiques des CM Couples;
- Par marginalisation, on obtient finalement:

$$E \left[X_{n+1} | y_1^N \right] = \sum_{r_{n+1}} E \left[X_{n+1} | r_{n+1}, y_1^N \right] p(r_{n+1} | y_1^N)$$

- Dans ce modèle, le processus (R, Y) est markovien, mais pas forcément R .

Lissage exact dans des modèles à sauts [Abbassi, 2012]

- On considère un modèle triplet $Z = (X_t, R_t, Y_t)_{t \in \mathcal{N}}$, tel que les X_t prennent leurs valeurs dans \mathbb{R}^{d_x} , les Y_t prennent leurs valeurs dans \mathbb{R}^{d_y} et $R_t \in \Omega$;
- Le modèle dynamique proposé est le suivant:

$R = (R_t)_{t \in \mathcal{N}}$ est une chaîne de Markov;

$$X_{t+1} = F_{t+1}(R_{t+1}, X_t) + H(R_{t+1}) W_{t+1}$$

$$Y_t = G_t(R_t, X_t) + J(R_t) V_t$$

- Les processus $W = (W_t)_{t \in \mathcal{N}}$ et $V = (V_t)_{t \in \mathcal{N}}$ sont centrés indépendants, et indépendants de X_0 ;



- $F_{t+1}(R_{t+1}, X_t)$, $G_t(R_t, X_t)$, $J(R_t)$ et $H(R_{t+1})$ sont des matrices variables dans le temps fonctions du processus X et du processus auxiliaire;
- Le triplet $Z = (X, R, Y)$ est un processus de Markov. De plus on montre que $S = (X, R)$ est markovien au vu de la récursivité;
- Le processus $X|R$ est donc markovien;
- Le problème du filtrage consistera à estimer à chaque instant la réalisation des deux variables cachées R_t et X_t à partir de $Y_0^t = (y_0, y_1, \dots, y_t)^T$;
- Ce calcul fait intervenir la densité de "filtrage", $p(s_0^t | y_0^t)$ qui s'exprime au moyen d'une récurrence:

$$\forall t \in \mathcal{N}, p(s_0^{t+1} | y_0^{t+1}) = \frac{p(z_{t+1} | z_t)}{p(y_{t+1} | y_0^t)} p(s_0^t | y_0^t)$$

- On retrouve le schéma du filtrage de Kalman "prédiction" + "filtrage".

Filtrage, prédiction

- Supposons qu'à l'instant t on connaisse la marginale $p(x_t|r_0^t, y_0^t)$ (filtrage à t) et la quantité $p(r_0^t|y_0^t)$;
- On peut alors en déduire $p(x_{t+1}|r_0^{t+1}, y_0^t)$ (prédiction à $t + 1$) ainsi que la quantité $p(y_{t+1}|r_0^{t+1}, y_0^t)$:

$$p(x_{t+1}|r_0^{t+1}, y_0^t) = \int p(x_{t+1}|x_t, r_{t+1})p(x_t|r_0^t, y_0^t)dx_t$$

$$p(y_{t+1}|r_0^{t+1}, y_0^t) = \int p(y_{t+1}|x_{t+1}, r_{t+1})p(x_{t+1}|r_0^{t+1}, y_0^t)dx_{t+1}$$

Filtrage, prédiction

- Et par récurrence on obtient la marginale $p(x_{t+1}|r_0^{t+1}, y_0^{t+1})$ (filtrage à $t + 1$);

$$p(x_{t+1}|r_0^{t+1}, y_0^{t+1}) = p(y_{t+1}|x_{t+1}, r_{t+1}) \frac{p(x_{t+1}|r_0^{t+1}, y_0^t)}{p(y_{t+1}|r_0^{t+1}, y_0^t)}$$

- Il faut calculer la quantité $p(r_0^{t+1}|y_0^{t+1})$ pour l'étape suivante:

$$p(r_0^{t+1}|y_0^{t+1}) \propto p(y_{t+1}|r_0^{t+1}, y_0^t) p(r_{t+1}|r_t) p(r_0^t|y_0^t)$$

- Il est possible de déduire $p(r_{t+1}|y_0^{t+1})$ par marginalisation:

$$p(r_{t+1}|y_0^{t+1}) = \sum_{r_0^t \in \Omega^{t+1}} p(r_0^{t+1}|y_0^{t+1})$$



Filtrage, prédiction

- Ainsi que le calcul de la quantité $p(x_{t+1}|y_0^{t+1})$:

$$p(x_{t+1}|y_0^{t+1}) = \sum_{r_0^{t+1} \in \Omega^{t+2}} p(r_0^{t+1}|y_0^{t+1})p(x_{t+1}|r_0^{t+1}, y_0^{t+1})$$

- En raison de l'aspect prohibitif en temps de calcul, le rafraîchissement de la quantité $p(r_0^t|y_0^t)$ ainsi que des variables d'intérêt x_t nécessitent par ex. une méthode de filtrage particulière;
- Dans le cas gaussien (bruits gaussiens) les marginales sont également gaussiennes, et les équations correspondent aux équations du filtre de Kalman et font intervenir récursivement les quantités "prédiction" $E[X_{t+1}|r_0^{t+1}, y_0^t]$ et "filtrage" $E[X_t|r_0^t, y_0^t]$.

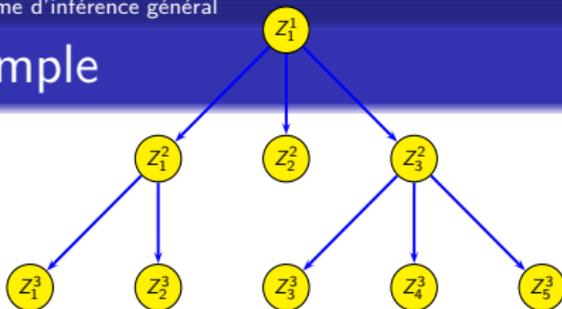
Hypothèses et notations [Lachantin, 2006]

- Soit un ensemble fini S et $\{S^1, S^2, \dots, S^N\}$ une partition de S ;
- On suppose que $\text{Card}S^1 = 1 \leq \text{Card}S^2 \leq \dots \leq \text{Card}S^N$;
- A chaque site $s \in S^n$ est associé un sous ensemble $s^+ \in S^{n+1}$ appelés "enfants" de s ;
- Les ensembles $\{s^+\}_{s \in S^n}$ forment une partition de S^{n+1} ;
- s^{++} désignent l'ensemble des descendants de s .
- De plus $\underline{s}^{++} = \{s\} \cup s^{++}$;
- En chaque site $s \in S \setminus S^1$, son unique parent est noté s^-
- La loi d'un signal multi-résolution Z s'exprime:

$$p(z) = p(z_1) \prod_{n=2}^N \prod_{s \in S^n} p(z_s | z_{S \setminus \underline{s}^{++}}) \quad (31)$$



Exemple



- La loi de Z s'écrit:

$$\begin{aligned}
 p(z) &= p(z_1^3, z_2^3, z_3^3, z_4^3, z_5^3 | z_1^1, z_1^2, z_2^2, z_3^2) p(z_1^2, z_2^2, z_3^2 | z_1^1) p(z_1^1) \\
 &= p(z_1^3 | z_1^1, z_1^2, z_2^2, z_3^2) \dots p(z_5^3 | z_1^1, z_1^2, z_2^2, z_3^2) p(z_1^2 | z_1^1) \dots p(z_3^2 | z_1^1) p(z_1^1)
 \end{aligned}$$

- En remarquant par exemple que:

$$z_3^2 \perp\!\!\!\perp z_1^2, z_2^2, z_1^3, z_2^3 | z_1^1$$

- Il vient:

$$p(z_3^2 | z_1^1) = p(z_3^2 | z_1^1, z_1^2, z_2^2, z_1^3, z_2^3)$$

- Les estimateurs du MAP et du MPM sont donnés par:

$$\forall s \in S, \hat{x}_s(y) = \sup_{x_s \in \Omega} p(x_s|y) \quad \hat{x}(y) = \sup_{x \in \Omega^{|S|}} p(x|y)$$

- L'estimation d'une réalisation de X nécessite le calcul des marginales a posteriori $p(x|y)$ et/ou $p(x_s|y)$ en chaque site;
- La loi a posteriori $p(x|y)$ peut être factorisée:

$$p(x|y) = p(x_1|y) \prod_{n=2}^N \prod_{s \in S^n} p(x_s|x_{S \setminus \underline{s}++}|y) \quad (32)$$

- Connaissant la loi initiale et les probabilités de transition a posteriori $p(x_s|x_{S \setminus \underline{s}++}|y)$, il est possible de calculer les marginales $p(x_s|y)$ selon une passe descendante:

$$p(x_s|y) = \sum_{x_{S \setminus \underline{s}++}} p(x_{S \setminus \underline{s}++}|y) \text{ où}$$

$$p(x_{S \setminus \underline{s}++}|y) = p(x_{S \setminus \underline{s}++}|y) p(x_s|x_{S \setminus \underline{s}++}, y)$$

Probabilités "rétrogrades"

- Soit la quantité:

$$\beta_{s,s^-}(x_{S \setminus \underline{s}^{++}}) = p(y_{\underline{s}^{++}} | z_{S \setminus \underline{s}^{++}})$$

- Relation entre probabilité rétrogrades et transitions:

$$p(x_s | x_{S \setminus \underline{s}^{++}}, y) = p(z_s | z_{S \setminus \underline{s}^{++}}) \frac{\prod_{t \in s^+} \beta_{t,s}(x_{S \setminus \underline{s}^{++}})}{\beta_{s,s^-}(x_{S \setminus \underline{s}^{++}})} \quad (33)$$

- Calcul récursif:

$$\beta_{s,s^-}(x_{S \setminus \underline{s}^{++}}) = \sum_{x_s \in \Omega} p(z_s | z_{S \setminus \underline{s}^{++}}) \prod_{t \in s^+} \beta_{t,s}(x_{S \setminus \underline{s}^{++}}) \quad (34)$$

Propriété de β dans son voisinage I

- Soit par exemple $t \in s^+$ un enfant du site s . Le site s est donc le parent de t . L'ensemble $S \setminus s^{++}$ correspond contient s et tous les sites sauf tous les descendants de s ;
- L'ensemble \underline{t}^{++} contient t et tous ses descendants. L'ensemble $s^{++} \setminus \underline{t}^{++}$ contient tous les descendants de s sauf ceux liés à la branche de t ;

$$\begin{aligned}
 \beta_{t,s}(x_{S \setminus \underline{t}^{++}}) &= p(y_{\underline{t}^{++}} | z_{S \setminus \underline{t}^{++}}) \\
 &= p(y_{\underline{t}^{++}} | z_{S \setminus s^{++}}, z_{s^{++} \setminus \underline{t}^{++}}) \\
 &= p(y_{\underline{t}^{++}} | z_{S \setminus s^{++}}) \\
 &= \beta_{t,s}(x_{S \setminus s^{++}})
 \end{aligned}$$

- Puisque $\forall t \in s^+$:

$$y_{\underline{t}^{++}} \perp\!\!\!\perp z_{s^{++} \setminus \underline{t}^{++}} | z_{S \setminus s^{++}}$$

Propriété de β dans son voisinage II

- Soit les deux seuls enfants de s , $t_1 \in s^+$ et $t_2 \in s^+$. C'est à dire $s^+ = \underline{t_1}^{++} \cup \underline{t_2}^{++}$. En utilisant les hypothèses d'indépendance conditionnelle, il est possible d'écrire:

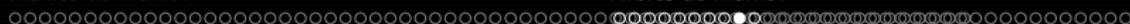
$$\begin{aligned} p(y_{s^{++}} | z_{S \setminus s^{++}}) &= p(y_{\underline{t_1}^{++}}, y_{\underline{t_2}^{++}} | z_{S \setminus s^{++}}) \\ &= p(y_{\underline{t_1}^{++}} | z_{S \setminus s^{++}}) p(y_{\underline{t_2}^{++}} | z_{S \setminus s^{++}}) \end{aligned}$$

- Ce résultat se généralise:

$$p(y_{s^{++}} | z_{S \setminus s^{++}}) = \prod_{t \in s^+} p(y_{\underline{t}^{++}} | z_{S \setminus s^{++}}) = \prod_{t \in s^+} \beta_{t,s}(x_{S \setminus s^{++}}) \quad (35)$$

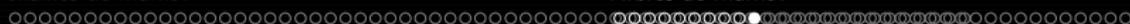
Probabilités "rétrogrades": démonstration de (34)

$$\begin{aligned}
 \beta_{s,s^-}(x_{S \setminus \underline{s}^{++}}) &= p(y_{\underline{s}^{++}} | z_{S \setminus \underline{s}^{++}}) \\
 &= p(y_s, y_{s^{++}} | z_{S \setminus \underline{s}^{++}}) \\
 &= \sum_{x_s \in \Omega} p(z_s, y_{s^{++}} | z_{S \setminus \underline{s}^{++}}) \\
 &= \sum_{x_s \in \Omega} p(y_{s^{++}} | z_{S \setminus \underline{s}^{++}}, z_s) p(z_s | z_{S \setminus \underline{s}^{++}}) \\
 &= \sum_{x_s \in \Omega} p(z_s | z_{S \setminus \underline{s}^{++}}) p(y_{s^{++}} | z_{S \setminus s^{++}}) \\
 &= \sum_{x_s \in \Omega} p(z_s | z_{S \setminus \underline{s}^{++}}) \prod_{t \in s^+} p(y_{\underline{t}^{++}} | z_{S \setminus s^{++}}) \\
 &= \sum_{x_s \in \Omega} p(z_s | z_{S \setminus \underline{s}^{++}}) \prod_{t \in s^+} \beta_{t,s}(x_{S \setminus s^{++}})
 \end{aligned}$$



Probabilités démonstration de (33)

$$\begin{aligned}
 p(x_s | x_{S \setminus \underline{s}++}, y) &= \frac{p(x_s, x_{S \setminus \underline{s}++}, y)}{p(x_{S \setminus \underline{s}++}, y)} \\
 &= \frac{p(x_{S \setminus s++}, y_{S \setminus s++}, y_{s++})}{p(x_{S \setminus \underline{s}++}, y_{S \setminus \underline{s}++}, y_{\underline{s}++})} \\
 &= \frac{p(z_{S \setminus s++}) p(y_{s++} | z_{S \setminus s++})}{p(z_{S \setminus \underline{s}++}) p(y_{\underline{s}++} | z_{S \setminus \underline{s}++})} \\
 &= p(z_s | z_{S \setminus \underline{s}++}) \frac{p(y_{s++} | z_{S \setminus s++})}{p(y_{\underline{s}++} | z_{S \setminus \underline{s}++})} \\
 &= p(z_s | z_{S \setminus \underline{s}++}) \frac{\prod_{t \in s^+} p(y_{t++} | z_{S \setminus s++})}{p(y_{\underline{s}++} | z_{S \setminus \underline{s}++})} \\
 &= p(z_s | z_{S \setminus \underline{s}++}) \frac{\prod_{t \in s^+} \beta_{t,s}(x_{S \setminus s++})}{\beta_{s,s^-}(x_{S \setminus \underline{s}++})}
 \end{aligned}$$



Algorithme général

Passé montante: $\forall n = N, \dots, 2 \forall s \in S^n$:

$$\beta_{s,s-}(x_{S \setminus \underline{s}++}) \propto \begin{cases} \sum_{x_s \in \Omega} p(z_s | z_{S \setminus \underline{s}++}) & \text{si } n = N \\ \sum_{x_s \in \Omega} p(z_s | z_{S \setminus \underline{s}++}) \prod_{t \in s+} \beta_{t,s}(x_{S \setminus s++}) & \end{cases}$$

$$p(x_s | x_{S \setminus \underline{s}++}, y) \propto \begin{cases} p(z_s | z_{S \setminus \underline{s}++}) & \text{si } n = N \\ p(z_s | z_{S \setminus \underline{s}++}) \prod_{t \in s+} \beta_{t,s}(x_{S \setminus s++}) & \end{cases}$$

Racine: $p(x_1 | y) \propto p(z_1) \prod_{t \in 1+} \beta_{t,1}(x_1)$

Passé descendante: $\forall n = N, \dots, 2 \forall s \in S^n$:

$$p(x_s | y) = \sum_{x_{S \setminus \underline{s}++}} p(x_{S \setminus s++} | y) \text{ où}$$

$$p(x_{S \setminus s++} | y) = p(x_{S \setminus \underline{s}++} | y) p(x_s | x_{S \setminus \underline{s}++}, y)$$

- On pose: $S^* = S \setminus (\underline{s}^{++} \cup \{s^-\})$ (en remarquant que $\underline{s}^{++} \cup \{s^-\} \neq \underline{s}^{-++}$ puisque le parent peut avoir plusieurs descendants);
- Le modèle classique consiste à établir les hypothèses suivantes:

$$X_s \perp\!\!\!\perp X_{S^*} | X_{s^-} \quad Y_s \perp\!\!\!\perp Z_{S \setminus \{s\}} | X_s$$

- La loi de Z est donnée par (31) avec:

$$\begin{aligned}
 p(z_s | z_{S \setminus \underline{s}^{++}}) &= p(z_s | z_{S^*}, z_{s^-}) \\
 &= p(x_s, y_s | x_{S^*}, y_{S^*}, x_{s^-}, y_{s^-}) \\
 &= p(x_s, y_s | x_{s^-}) \\
 &= p(x_s | x_{s^-}) p(y_s | x_s)
 \end{aligned}$$

- En observant que $Y_{\underline{s}^{++}} \perp\!\!\!\perp (Y_{s^-}, Z_{s^*}) | X_{s^-}$ la probabilité rétrograde s'écrit:

$$\begin{aligned}
 \beta_{s,s^-}(x_{S \setminus \underline{s}^{++}}) &= p(y_{\underline{s}^{++}} | z_{S \setminus \underline{s}^{++}}) \\
 &= p(y_{\underline{s}^{++}} | z_{s^*}, x_{s^-}, y_{s^-}) \\
 &= p(y_{\underline{s}^{++}} | x_{s^-}) \\
 &= \beta_{s,s^-}(x_{s^-})
 \end{aligned}$$

- De même $\forall t \in s^+, \beta_{t,s}(x_{S \setminus s^{++}}) = \beta_{t,s}(x_s)$

Markoviannité de X *a posteriori*

- D'après (33), et ce qui précède, la loi de transition de X *a posteriori* est donc donnée par :

$$p(x_s | x_{S \setminus \underline{s}^{++}}, y) = p(x_s | x_{s-}) p(y_s | x_s) \frac{\prod_{t \in s^+} \beta_{t,s}(x_s)}{\beta_{s,s-}(x_{s-})} = p(x_s | x_{s-}, y) \quad (36)$$

- On en déduit que la variable $X|Y = y$ est markovienne.
- La connaissance de la loi initiale et des lois de transition *a posteriori* (passe montante) permet la simulation de réalisation $X|y = y$ et l'estimation des hyper-paramètres (EM, SEM, ICE etc.);
- Segmentation suite à la passe descendante.



Sommaire

- 1 chaînes de Markov
 - Algorithme d'inférence général
 - Chaînes de Markov Cachées à bruit indépendant (CMC-BI)
 - Chaîne de Markov couples (CMCo)
 - Chaînes couples partiellement de Markov (CCPM)
 - Chaînes de Markov triplets (CMT)
 - Exemples
- 2 Arbres de Markov
 - Algorithme d'inférence général
 - Arbres de Markov caché à bruit indépendant (AMC-BI)
 - **Arbre de Markov Couple (AMCo)**
 - Arbre Couple partiellement de Markov (ACPM)
 - Exemples
- 3 Conclusion
 - Conclusion

- Ce modèle consiste à poser directement la markoviannité en échelle du processus Z :

$$Z_s \perp\!\!\!\perp Z_{S^*} | Z_{s-}$$

- La loi de Z est donnée par (31) avec:

$$\begin{aligned} p(z_s | z_{S \setminus \underline{s}^{++}}) &= p(z_s | z_{S^*}, z_{s-}) \\ &= p(z_s | z_{s-}) \end{aligned}$$

- Les hypothèses d'indépendance donnent:

$$\begin{aligned}
 \beta_{s,s^-}(x_{S \setminus \underline{s}^{++}}) &= p(y_{\underline{s}^{++}} | z_{S \setminus \underline{s}^{++}}) \\
 &= p(y_{\underline{s}^{++}} | z_{S^*}, z_{s^-}) \\
 &= p(y_{\underline{s}^{++}} | z_{s^-}) \\
 &= \beta_{s,s^-}(x_{s^-})
 \end{aligned}$$

- De même $\forall t \in s^+, \beta_{t,s}(x_{S \setminus s^{++}}) = \beta_{t,s}(x_s)$
- Expression des lois de transition *a posteriori*:

$$p(x_s | x_{S \setminus \underline{s}^{++}}, y) = p(z_s | z_{s^-}) \frac{\prod_{t \in s^+} \beta_{t,s}(x_s)}{\beta_{s,s^-}(x_{s^-})} = p(x_s | x_{s^-}, y) \quad (37)$$

- $X|Y = y$ (resp. $Y|X = x$) est markovien. Ainsi, des approches analogues aux chaînes couples sont possibles.

- Hypothèses d'indépendance conditionnelle:

$$Z_s \perp\!\!\!\perp X_{S^*} | Z_{s^-}, Y_{S^*}$$

- La loi de Z est donnée par (31) avec:

$$\begin{aligned} p(z_s | z_{S \setminus \underline{s}^{++}}) &= p(z_s | x_{S^*}, y_{S^*}, x_{s^-}, y_{s^-}) \\ &= p(z_s | x_{s^-}, y_{S \setminus \underline{s}^{++}}) \end{aligned}$$

- Probabilité “rétrograde” :

$$\begin{aligned}
 \beta_{s,s^-}(x_{S \setminus \underline{s}^{++}}) &= p(y_{\underline{s}^{++}} | z_{S \setminus \underline{s}^{++}}) \\
 &= p(y_{\underline{s}^{++}} | z_{S^*}, z_{s^-}) \\
 &= p(y_{\underline{s}^{++}} | x_{S^*}, y_{S^*}, z_{s^-}) \\
 &= p(y_{\underline{s}^{++}} | x_{s^-}, y_{S \setminus \underline{s}^{++}}) \\
 &= \beta_{s,s^-}(x_{s^-}, y_{S \setminus \underline{s}^{++}})
 \end{aligned}$$

- Expression des lois de transition *a posteriori*:

$$p(x_s | x_{S \setminus \underline{s}^{++}}, y) = p(z_s | x_{s^-}, y_{S \setminus \underline{s}^{++}}) \frac{\prod_{t \in s^+} \beta_{t,s}(x_s, y_{S \setminus \underline{t}^{++}})}{\beta_{s,s^-}(x_{s^-}, y_{S \setminus \underline{s}^{++}})} = p(x_s | x_{s^-}, y)$$

- $X|Y = y$ est markovien. Ainsi, des approches analogues aux CCPM sont possibles.



Sommaire

- 1 chaînes de Markov
 - Algorithme d'inférence général
 - Chaînes de Markov Cachées à bruit indépendant (CMC-BI)
 - Chaîne de Markov couples (CMCo)
 - Chaînes couples partiellement de Markov (CCPM)
 - Chaînes de Markov triplets (CMT)
 - Exemples
- 2 Arbres de Markov
 - Algorithme d'inférence général
 - Arbres de Markov caché à bruit indépendant (AMC-BI)
 - Arbre de Markov Couple (AMCo)
 - Arbre Couple partiellement de Markov (ACPM)
 - Exemples
- 3 Conclusion
 - Conclusion

Lissage avec arbre triplet à sauts markoviens

- On considère un processus triplet
 $T = (X, R, Y) = (X_s, R_s, Y_s)_{s \in S}$, les X_s sont à valeurs dans \mathbb{R}^q , les Y_s à valeurs dans \mathbb{R}^m et les V.A. R_s à valeurs dans un ensemble fini Λ^S ;
- (X, R) est caché et Y est observé;
- Le problème du lissage consiste à calculer $E[X_s | Y]$;
- On considère un arbre “mono-parental” dont les enfants sont indépendants conditionnellement au parent;
- Loi d'un arbre de markov triplet:

$$p(t) = p(t_1) \prod_{s \in S \setminus s_1} p(t_s | t_{s-})$$

- Si l'on reprend le modèle classique précédent:

$R = (R_n)_{n \in \mathcal{N}}$ est un arbre de Markov;

$$X_s = F_s(R_s) X_{s^-} + G_s(R_s) W_s$$

$$Y_s = H_s(R_s) X_s + J_s(R_s) V_s$$

- Avec W_s et V_s sont indépendantes et gaussiennes;
- Les matrices $F_s(R_s)$, $G_s(R_s)$, $H_s(R_s)$ et $J_s(R_s)$ dépendent de R_s ;
- Le calcul de $E[X_s|Y]$ avec une complexité linéaire en $|S|$ est prohibitif;

- Soit dans ces conditions (processus R caché) le modèle suivant:

(R, Y) est un arbre de Markov;

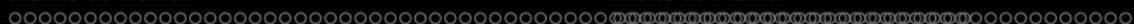
$$X_s = F_s(R_s, Y_s) X_{s-} + G_s(R_s, Y_s) W_s \quad (38)$$

- Dans ce cas, le processus $T = (X, R, Y)$ est markovien et le calcul du lissage n'est pas prohibitif;
- On calcule récursivement $E[X_s | r_s, Y]$ en fonction de $E[X_{s-} | r_{s-}, Y]$:

$$E[X_s | r_s, y_1^N] = F_s(r_s, y_s) \sum_{r_{s-}} E[X_{s-} | r_{s-}, y_1^N] p(r_{s-} | r_s, y_1^N)$$

- Les marginales et lois de transition a posteriori peuvent être calculées par traitement de l'arbre couple de Markov;
- Finalement, l'on déduit:

$$E[X_s | y] = \sum_{r_s} E[X_s | r_s, y] p(r_s | y)$$



- Lien entre les modèles triplet et la théorie de l'évidence [Pieczynski, 2003, 2012];
- Bruits non gaussiens corrélés [Lanchantin, 2011];
- Chaînes triplets et bruit à mémoire longue [Lanchantin, 2008], [Abbassi, 2012];
- Lissage exact avec modèles couples, modèles triplets [Desbouvries, 2006, 2013];
- Modèle de chaîne semi-Markov (marginale d'une chaîne couple) non stationnaires [Lapuyade, 2010, 2011] ;
- Modèles par champs de Markov [Benboudjema, 2007];
- Modèle général de chaîne de Markov couple non stationnaire [Ben Mabrouk, 2011] où la chaîne couple marginale de la chaîne triplet n'est pas forcément stationnaire;
- Variantes avec variables mixtes X est discret et U est continu [Ben Mabrouk, 2011];