TRIPLET MARKOV MODELS
Wojciech Pieczynski, Telecom Sudparis, October 2016

A. Hidden discrete chains
1. Hidden and Pairwise Markov;
2. Triplet Markov;
3. Parameter estimation;
4. Further extensions.

B. Hidden continuous chains
1. Classic Kalman filter in Gaussian HMMs;

C. Switching conditionally linear models
1. Classic extension: Conditionally Gaussian Linear State-Space Model (CGLSSM);
2. Approximating any non linear Markov system with CGOMSM;
3. General “conditionally Markov switching hidden linear models” (CMSHLMs).

D. Triplet Markov fields and image segmentation
1. Evidential Markov models;
2. Image segmentation
A. Hidden discrete chains

1. Hidden, pairwise, and triplet Markov

Let $X = (X_1, ..., X_n)$, $Y = (Y_1, ..., Y_n)$, with $X_i \rightarrow \Omega = \{1, ..., K\}$ and $Y_i \rightarrow R$. Let $Z = (Z_1, ..., Z_n)$, with $Z_i = (X_i, Y_i)$.

Hidden Markov (HMM):

$$p(x, y) = \frac{p(x_1) p(x_2|x_1) p(x_n|x_{n-1}) p(y_1|x_1) p(y_2|x_2) ... p(y_n|x_n)}{p(x) p(y|x)}$$

Pairwise Markov (PMM):

$$p(z) = p(z_1) p(z_2|z_1) ... p(z_n|z_{n-1})$$
As \( p(z_{i+1}|z_i) = p(x_{i+1}|x_i, y_i)p(y_{i+1}|x_{i+1}, x_i, y_i) \),

\[
p(x_{i+1}|x_i, y_i) = p(x_{i+1}|x_i), \text{ and}
\]

\[
p(y_{i+1}|x_{i+1}, x_i, y_i) = p(y_{i+1}|x_{i+1}),
\]

which can be seen as an **useless – and sometimes strong - approximation**, at least for Bayesian segmentation.
For example $p(x_i|y)$ is computed in PMM with the same complexity as in HMM. Setting

$$\alpha(x_i) = p(x_i, y_1, \ldots, y_i) \;; \beta(x_i) = p(y_{i+1}, \ldots, y_n|x_i, y_i),$$

we have

$$\alpha(x_1) = p(z_1), \alpha(x_{i+1}) = \sum_{x_i \in \Omega} \alpha(x_i) p(x_{i+1}, y_{i+1}|x_i, y_i) \frac{p(z_{i+1}|z_i)}{p(z_{i+1}|z_i)} \text{ for } 2 \leq i \leq n;$$

and

$$\beta(x_n) = 1, \beta(x_i) = \sum_{x_{i+1} \in \Omega} \beta(x_{i+1}) p(x_{i+1}, y_{i+1}|x_i, y_i) \frac{p(z_{i+1}|z_i)}{p(z_{i+1}|z_i)} \text{ for } 1 \leq i \leq n - 1,$$

which are the same as in HMMs with $p(z_{i+1}|z_i)$ instead of $p(x_{i+1}|x_i)p(y_{i+1}|x_{i+1})$. Then $p(x_i, y) = \alpha_i(x_i)\beta_i(x_i)$, which gives $p(x_i|y)$. 


Dependence graphs are

HMM

PMM
We have the following NS conditions showing that Markovianity of $X$ can be a constraint:

**Proposition 1**

In stationary revertible PMM $Z = (X, Y)$ the following conditions:

(i) $X$ is Markov;

(ii) $p(y_i|x_i, x_{i-1}) = p(y_i|x_i)$ for each $2 \leq i \leq n$;

(iii) $p(y_i|x) = p(y_i|x_i)$ for each $1 \leq i \leq n$,

are equivalents.
Remark 1.

One can have «hidden» Markov with correlated noise (HMM-CN), and “pairwise” (with X non Markov) with independent noise (PMM-IN).
Remark 2.

In the context of Proposition 1, in PMM with non-Markov $X$, $p(y_i|x)$ depends on all $x_1, \ldots, x_n$ and the marginal noise distribution $p(y_i|x_i)$ is a rich mixture:
Do PMMs work better than HMMs in real situations?

Examples in image segmentation

Bi-dimensional set of pixels is converted into monodimensional sequence via the Hilbert-Peano scan:
Gaussian correlated noise is obtained with mobile mean; Parameters estimated with ICE

<table>
<thead>
<tr>
<th>(X = x)</th>
<th>(Y = y)</th>
<th>HMM</th>
<th>Error</th>
<th>PMM</th>
<th>Error</th>
<th>« True » Par</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td>13.3%</td>
<td><img src="image4.png" alt="Image" /></td>
<td>7.7%</td>
<td>7.5%</td>
</tr>
<tr>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
<td><img src="image7.png" alt="Image" /></td>
<td>17.4%</td>
<td><img src="image8.png" alt="Image" /></td>
<td>9.1%</td>
<td>7.1%</td>
</tr>
</tbody>
</table>

One notices good efficiency of ICE
2. Triplets Markov

Let \((X, Y)\) as above, and \(U = (U_1, \ldots, U_n)\), with \(U_i \rightarrow \Lambda = \{1, \ldots, M\}\). 

\(U\) can have real signification or not

Let \(T = (X, U, Y)\) be Markov. Then:

(i) \(T = (V, Y)\), with \(V = (X, U)\), is a PMM. Thus posterior margins

\[
p(x_i | y) = \sum_{u_i} p(x_i, u_i | y)
\]

are computable;

(ii) \((X, Y)\) is not necessarily Markov: TMM are strictly more general than PMMs.
Examples

1. TMM $T = (X, U, Y)$ can be quite simple: $U$ is a Markov chain and
\[
p(x, y|u) = \prod_{i=1}^{n} p(x_i|u_i)p(y_i|u_i).
\]

None of chains $X$, $Y$, $(X, Y)$ is Markov in general.
2. Particular Hidden Semi-Markov Models (HSMMs) are TMCs.

3. As $T = (V,Y)$, with $V = (X,U)$, is a PMM, parameter estimation methods are the same in PMMs and TMMs.

4. Let $V = (X,U)$ be Markov, and $p(y|u,x) = \prod_{i=1}^{n} p(y_i|x_i)$.

There are $M \times K$ classes, but only $K$ Gaussian densities
$U$ models three « textures », or three « stationarities »; two Gaussian densities $N(0,1), N(2,1)$. Parameters estimated with EM

<table>
<thead>
<tr>
<th>$X = x$</th>
<th>$Y = y$</th>
</tr>
</thead>
</table>

| HMM : $\tau = 7.2\%$ | TMM : $\tau = 3.5\%$ | $\hat{U} = \hat{u}$ |
3. Parameter estimation

3.1 EM and ICE

Let $T = (X, U, Y) = (V, Y)$ be a stationary TMM, with $p(v, y)$ defined by $p_\theta(v_1, v_2, y_1, y_2)$. Two general iterative methods:

- find initial $\theta^0$;

- EM: $\theta^{q+1} = \arg\max_{\theta} E[p_\theta(V, Y) \mid Y = y, \theta^q]$, 

- ICE: $\theta^{q+1} = E[\hat{\theta}(V, Y) \mid Y = y, \theta^q]$.

EM and ICE work well in Gaussian case.
### 3.2 Generalized ICE

Let \( p_\theta(v_1, v_2, y_1, y_2) = p_\theta(v_1, v_2)p(y_1, y_2 \mid v_1, v_2) \). In the general case

\[
p_{ij}^\theta(y_1, y_2) = p_\theta(y_1, y_2 \mid v_1 = i, v_2 = j)
\]

are defined by:

(i) copulas \( c_{ij}^\theta(y_1, y_2) \) (probability density on \([0,1]^2\) such that margins are identity), and

(ii) margins \( p_{ij}^\theta(y_1), p_{ij}^\theta(y_2) \).

In fact, setting \( F_{ij}^\theta(y_1), F_{ij}^\theta(y_2) \) cumulative distribution functions, we have (Sklar theorem):
$$p_{\theta}^{ij}(y_1, y_2) = p_{\theta}^{ij}(y_1) p_{\theta}^{ij}(y_2) c_{\theta}^{ij}(F_{\theta}^{ij}(y_1), F_{\theta}^{ij}(y_2))$$

Recent (2016) “generalized” ICE allows to search, from $Y = y$:

- distribution $p(\nu_1, \nu_2, y_1, y_2)$;

- copulas $c_{\theta}^{ij}$ and margins $p_{\theta}^{ij}$, knowing that each $c_{\theta}^{ij}$ belongs to one among families $C_{ij} = \{C_{ij,1},...,C_{ij,m_3(i,j)}\}$ of copulas, and $p_{\theta}^{ij}$ belongs to one among families $H_{ij} = \{F_{ij,1},...,F_{ij,m_1(i,j)}\}$ of margins.
4. Further extensions

4. 1. Dempster-Shafer theory of evidence

Fuzzy measure :

(i) \( M(\Omega) = 1, \ M(\emptyset) = 0; \)

(ii) \( A \subset B \Rightarrow M(A) \leq M(B), \)

can be of interest in sensors fusion (image processing). TMMs allow one to consider some fuzzy measures in Markov context.
4. 2. Triplet Partially Markov Models

Let $T = (V, Y)$ with $V = (V_1, ..., V_n)$ discrete and $Y = (Y_1, ..., Y_n)$ continuous. $T = (V, Y)$ is a Pairwise Partially Markov Model (PPMM) if

$$p(v, y) = p(v_1, y_1)p(v_2, y_2|v_1, y_1)p(v_3, y_3|v_2, y_1, y_2)\cdots p(v_n, y_n|v_{n-1}, y_1, ..., y_{n-1})$$

All processing remain feasible.

In particular, $T = (X, U, Y)$ is workable if $T = (V, Y)$, with $V = (X, U)$, a PPMM ($T = (X, U, Y)$ is called TPMM).
B. Continuous hidden chain

1. Kalman filter in Gaussian HMMs, PMMs, and TMMs.

$(X, Y)$ is a Gaussian HMM if

\[
X_{n+1} = F_n X_n + G_n W_n;
\]

\[
Y_n = H_n X_n + J_n N_n,
\]

with $(W_n)$, $(N_n)$ white Gaussian noises independent from each other.
Computing $Y_{n+1}$ from $X_n$ gives: $Y_{n+1} = H_{n+1}X_{n+1} + J_{n+1}N_{n+1} = H_{n+1}(F_nX_n + G_nW_n) + J_{n+1}N_{n+1} = K_nX_n + L_nW_n + J_{n+1}N_{n+1}$.
\[
\begin{bmatrix}
X_{n+1} \\
Y_{n+1} \\
Z_{n+1}
\end{bmatrix} =
\begin{bmatrix}
F_n & 0 & X_n \\
K_n & 0 & Y_n \\
A_n & 0 & Z_n
\end{bmatrix} +
\begin{bmatrix}
G_n & 0 & W_n \\
L_n & J_{n+1} & N_{n+1}
\end{bmatrix}
\]

Kalman filter is still workable in PMMs:

\[
\begin{bmatrix}
X_{n+1} \\
Y_{n+1}
\end{bmatrix} =
\begin{bmatrix}
A_{n+1} & B_{n+1} \\
C_{n+1} & D_{n+1}
\end{bmatrix}
\begin{bmatrix}
X_n \\
Y_n
\end{bmatrix} +
\begin{bmatrix}
E_{n+1} & A_{n+1} \\
L_{n+1} & J_{n+1}
\end{bmatrix}
\begin{bmatrix}
W_{n+1} \\
N_{n+1}
\end{bmatrix}
\]
Finally, PMMs can be extended to Gaussian TMMs

Optimal fast filter is workable while \((X, Y)\) is not Markov. This opens ways for the use of complex \(U\) leading to complex non Markovian \((X, Y)\).
C. Switching conditionally linear models

1. Classic extension: Conditionally Gaussian Linear State-Space Model (CGLSSM)

Let $R = (R_1, \ldots, R_n)$ be a Markov process of «switches», with $R_i \rightarrow \Lambda = \{1, \ldots, M\}$.

CGLSSM verifies:

$$X_{n+1} = F_n(R_{n+1})X_n + G_n(R_{n+1})W_n;$$

$$Y_n = H_n(R_n)X_n + J_n(R_n)N_n,$$

which is a “natural” extension of HMM to the switching case.
Fast filtering is not workable; in particular because $p(r_i|y_1,\ldots,y_i)$ is not computable recursively. Different approximations like particle filtering
CGLSSM above can be extended to “Conditionally Gaussian Markov Switching Model” (CGMSM):

- triplet \((X, R, Y)\) is Markov, and \(p(r_{i+1}|x_i, r_i, y_i) = p(r_{i+1}|r_i)\) (then \(R = (R_1,..., R_n)\) is Markov);

\[
\begin{bmatrix}
X_{n+1} \\
Y_{n+1}
\end{bmatrix} = \begin{bmatrix}
A_1(R_n^{n+1}) & A_2(R_n^{n+1}) \\
A_3(R_n^{n+1}) & A_4(R_n^{n+1})
\end{bmatrix} \begin{bmatrix}
X_n \\
Y_n
\end{bmatrix} + \begin{bmatrix}
B_1(R_n^{n+1}) & B_2(R_n^{n+1}) \\
B_3(R_n^{n+1}) & B_4(R_n^{n+1})
\end{bmatrix} \begin{bmatrix}
U_{n+1} \\
V_{n+1}
\end{bmatrix}
\]

Of course, fast filtering is not workable any more …
However, in particular “Conditionally Gaussian Observed Markov Switching Model” (CGOMSM) obtained with $A_3(r_i^{i+1}) = 0$ (for each $i = 1, \ldots, n - 1$ and $r_i^{i+1} \in \Lambda^2$)) fast filtering is workable.

Main difference between CGLSSM and CGOMSM is:

(i) in CGLSSM $(X, R)$ is Markov and $(R, Y)$ is not;

(ii) in CGOMSM $(R, Y)$ is Markov and $(X, R)$ is not.
Conditionally Gaussian Markov Switching Model” (CGMSM)

Classic CGLSSM

Recent CGOMSM
Classic CGLSSM

Recent CGOMSM
Remark 3

The classic CGLSSM distribution is given by $p(x_1, r_1, y_1)$ and the transitions $p(r_{i+1}|r_i)$, $p(x_{i+1}|r_{i+1}, r_i, x_i)$, $p(y_{i+1}|r_{i+1}, x_{i+1})$. One can show that there exists numerous CGOMSMs having the same transitions.
2. Approximating any non linear Markov system with CGOMSM

Let \((X,Y)\) be stationary Markov with the distribution \(p(x, y)\) given by \(p(x_1, y_1)\) and the recursions:

\[
(X_{i+1}, Y_{i+1}) = H(X_i, Y_i, W_i),
\]

with \(W = (W_1, ..., W_n)\) independent variables.

We assume that it is possible to sample realizations of \((X,Y)\).
Distribution $p(x, y)$ is defined by $p(x_1, y_1, x_2, y_2)$. Let

$$p(x_1, y_1, x_2, y_2) \approx \sum_{(i, j) \in \Lambda^2} \alpha_{ij} p_{ij}(x_1, y_1, x_2, y_2)$$

be an approximation with Gaussian mixture, which can be seen as the marginal distribution of $p(x_1, r_1, y_1, x_2, r_2, y_2)$:

$$p(r_1 = i, r_2 = j) p(x_1, y_1, x_2, y_2 | r_1 = i, r_2 = j)$$

$$\alpha_{ij} p_{ij}(x_1, y_1, x_2, y_2)$$
Taking \( p(x_1, r_1, y_1, x_2, r_2, y_2) \) such that triplet Markov defined with is a CGOMSM, one can the estimate it from data sampled with the system.

Indeed, CGOMSM defined with \( p(r_1, r_2)p(x_1, y_1, x_2, y_2|r_1, r_2) \) can be seen as a hidden Markov \((R, Z)\), with \( Z = (X, Y) \), and thus it can be estimated from \( Z_1, Z_2, \ldots, Z_N \) sampled with the system via some method.

Only one needs is to be able to sample realizations of the system.
3. Fast exact filtering in general “conditionally Markov switching hidden linear models” (CMSHLMs)

CGOMSMs are particular following CMSHLMs in which we will detail the fast optimal filter. \((X, R, Y)\) is a CMSHLM if

(i) \((X, R, Y)\) is Markov with

\[
p(r_{i+1}, y_{i+1}|x_{i+1}, r_{i+1}, y_{i+1}) = p(r_{i+1}, y_{i+1}|r_{i+1}, y_{i+1})
\] (3.1)

((\(R, Y\) is then Markov);

(ii) \(X_{n+1} = A_{n+1}(R_n, Y_n, R_{n+1}, Y_{n+1})X_n + H_{n+1}(R_n, Y_n, R_{n+1}, Y_{n+1})\), with

\[
E[H_{n+1}|R_n, Y_n, R_{n+1}, Y_{n+1}] = M_{n+1}(R_n, R_{n+1})
\]
We have to compute $E[X_{n+1}\mid y_1^{n+1}]$ from $E[X_n\mid y_1^n]$ and $y_{n+1}$; it is sufficient to compute $E[X_{n+1}\mid r_{n+1}, y_1^{n+1}]$ and $p(r_{n+1}\mid y_1^{n+1})$ from $E[X_n\mid r_n, y_1^n]$, $p(r_n\mid y_1^n)$, and $y_{n+1}$.

Taking conditional expectation of (ii):

$$E[X_{n+1}\mid r_n, r_{n+1}, y_1^{n+1}] = A_{n+1}(r_n, y_n, r_{n+1}, y_{n+1})E[X_n\mid r_n, r_{n+1}, y_1^{n+1}] + M_{n+1}(r_n, r_{n+1})$$
The crux point is

\[ E[X_n | r_n, r_{n+1}, y_1^{n+1}] = E[X_n | r_n, y_1^n], \text{ indeed:} \]

![Diagram of the crux point]
Besides $E[X_{n+1} | r_{n+1}, y_1^{n+1}] = \sum_{r_n} E[X_{n+1} | r_{n}, r_{n+1}, y_1^{n+1}] p(r_n | r_{n+1}, y_1^{n+1})$, thus

$$E[X_{n+1} | r_{n+1}, y_1^{n+1}] = \sum_{r_n} [A_{n+1}(r_n, y_n, r_{n+1}, y_{n+1}) E[X_n | r_n, y_1^n] + M_{n+1}(r_n, r_{n+1})] p(r_n | r_{n+1}, y_1^{n+1})$$

As $p(r_n, r_{n+1} | y_1^{n+1})$ is computable with ($(R,Y)$ is Markov):

$$p(r_n, r_{n+1} | y_1^{n+1}) = \frac{p(r_n, r_{n+1}, y_{n+1} | y_1^n)}{p(y_{n+1} | y_1^n)} = \frac{p(r_{n+1}, y_{n+1} | r_n, y_n)}{p(y_{n+1} | y_1^n)} p(r_n | y_1^n)$$

with
\[ p(y_{n+1}|y_1^n) = \sum_{r_n, r_{n+1}} p(r_n, r_{n+1}, y_{n+1}|y_1^n) = \sum_{r_n, r_{n+1}} p(r_n|y_1^n)p(r_{n+1}, y_{n+1}|r_n, y_n) \]

which gives \( p(r_{n+1}|y_1^{n+1}) \) and \( p(r_n|r_{n+1}, y_1^{n+1}) \).

Finally \( E[X_{n+1}|r_{n+1}, y_1^{n+1}] \) and \( p(r_{n+1}|y_1^{n+1}) \) are computable from

\[ E[X_n|r_n, y_1^n] \text{ et } p(r_n|y_1^n) \] with a finite, independent from \( n \), number of operations.
Remark 4

Rich possibilities of extensions: the Markov chain \((R, Y)\) in TMM \((X, R, Y)\) can be extended to different models discussed in part A: one can add a fourth chain \(U\) such that the riplet \((R, U, Y)\) is Markov, and then \((X, R, U, Y)\) remains workable.
D. Triplet Markov fields and image segmentation

1. Evidential hidden Markov models

\[ p(x, m, u, y) \propto \]

\[ \exp\left[-\sum_{c \in C} \psi_c(m_c) - \sum_{c \in C} \phi_c(u_c) - \sum_{s \in S} \eta_s(x_s, u_s) + \sum_{s \in S} \log(p(y_s|x_s, u_s))\right] \]

\( X \): Class field;

\( U \): Auxiliary field, models Gaussian mixture and noise correlation;

\( M \): Evidential field, models “more” or “les” spatial dependence.
2. Unsupervised image segmentation

- Noisy Image
- Class image
- "pixel by pixel" method
- HMF based method
- HEMF based method
Noisy image

Class image

“Blind” segmentation  HMFs segmentation  EHMFs segmentation
Optical image

Histogram

“pixel by pixel”  HMF  HEMF
Conclusions

1. Main idea: from mathematical viewpoint (and often physical one) X and Y play symmetrical roles;

2. Same fast processing in very general models (the couple (X, Y) may even not be Markov);

3. Good efficiency of ICE: unsupervised segmentation;

4. In discrete case error ratio divided by two; other evaluations in continuous cases in progress;

5. Further extensions: Dempster-Shafer theory, long-memory noises, Markov fields, Markov networks, … all possibly applied to “big data”.