Multichain hidden Markov and semi-Markov processes with applications

Jean-Baptiste Durand[†], Hanna Bacave[‡], Alain Franc[⋆],
Sandra Plancade[‡], Nathalie Peyrard[‡] and Régis Sabbadin[‡]

Processus markoviens, semi-markoviens et leurs applications
5-7 juin 2023 - Montpellier

Project supported by ANR HSMM-INCA

† Cirad, UMR AMAP, Montpellier ‡ INRAE, Unité MIA, Toulouse,
⋆ INRAE UMR BioGeCo, Bordeaux

Multichain hidden Markov models (HMMs): motivation, definition and illustrations

Parameter estimation: impact of couplings on complexity

Perspectives et conclusions

Motivation for Multichain HMMs

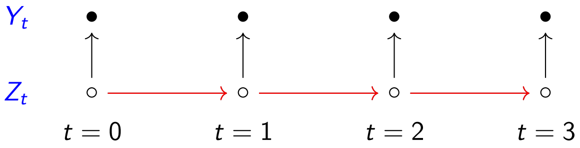
- HMM: classical statistical model for analysis of discrete-time latent signals (ecology, medicine, natural hazards, video analysis, ...)
- In many applications: more than one hidden chain; interaction between chains
  - metapopulation dynamics on a network of patches
  - disease spread over a network of hosts
  - earthquake activities in neighbour seismic areas
- Questions
  - How to formalize the concept of 'Multichain' HMM?
  - How is inference complexity impacted by additional chains?

An HMM is a directed graphical model

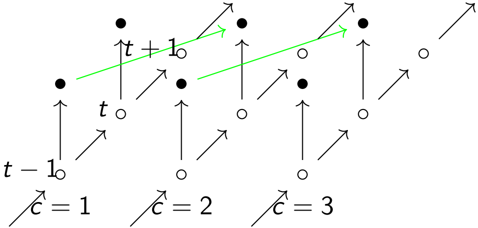
- Let us define
  $Z_{0:T} = \{Z_0, \ldots, Z_T\}$: the sequence of hidden states
  $Y_{0:T} = \{Y_0, \ldots, Y_T\}$: the sequence of observed states
- Joint distribution

$$\begin{aligned}\mathbb{P}(Z_{0:T} = z_{0:T}, Y_{0:T} = y_{0:T}) \quad &= \quad \mathbb{P}(Z_0 = z_0)\mathbb{P}(Y_0 = y_0 \mid Z_0 = z_0)\ldots \\ &\times \prod_{t=1}^{T} \mathbb{P}(Z_t = z_t | Z_{t-1} = z_{t-1})\mathbb{P}(Y_t = y_t | Z_t = z_t).\end{aligned}$$

- Graphical representation of conditional independencies



Unformal definition of Multichain HMMs



- The set of variables at t separate those at $t-1$ and those at $t+1$
- The HMM graph is a subgraph of the whole graph for each c
- There may be other edges between chains joining variables at $t-1$ and t

• Remarks

- This definition does not imply that $(Z^c, Y^c) = ((Z_t^c)_t, (Y_t^c)_t)$ is a HMM

- This definition includes Factorial HMM (FHMM, Ghahramani and Jordan, 1997) and Coupled HMM (CHMM, Brand et al., 1997; Wainwright and Jordan, 2008)

Proposed definition of Multichain HMMs

Now C couples of sequences $(Z^c, Y^c) = ((Z_t^c)_t, (Y_t^c)_t)$ for $1 \leq c \leq C$, and $\boldsymbol{Z} = (Z^1, \ldots, Z^C)$, $\boldsymbol{Y} = (Y^1, \ldots, Y^C)$ and $\boldsymbol{X} = ((Z^1, Y^1), \ldots, (Z^C, Y^C))$.

Definition

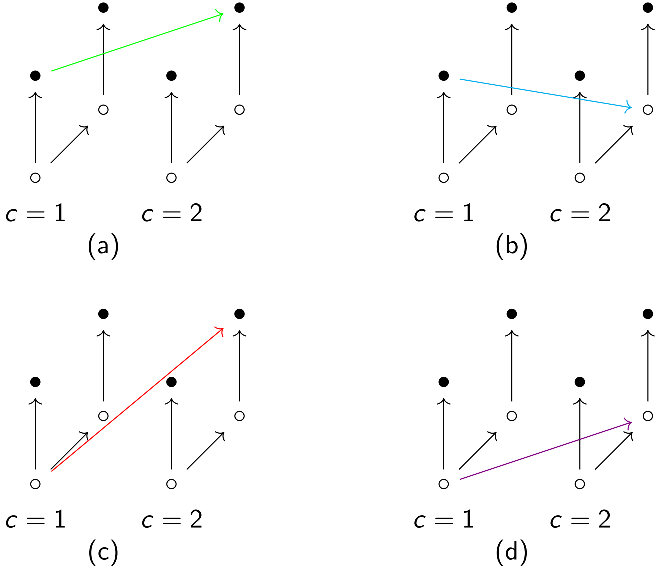
We say that the distribution of $(\boldsymbol{Z}, \boldsymbol{Y})$ is a multichain HMM if

(i) the joint distribution of $\boldsymbol{X}$ satisfies the Markov property,

(ii) in the graphical representation of the conditional independencies, there is an edge from Z_t^c towards Y_t^c and an edge from Z_t^c towards Z_{t+1}^c,

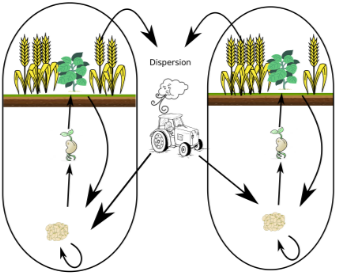
(iii) there may exist other edges between variables at time t and variables at time $t+1$.

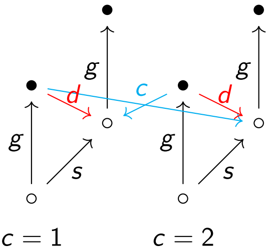
Typology of elementary couplings



Examples from the literature with applications

Structure (b) : Weeds Le Coz et al., 2019





where s is seed survival, g germination, c colonization and d seed production.

Others applications

(d) Recognition of human movements (Brand et al., 1997)
(d) Coupled HMMs: spread of infection (Touloupou et al., 2020)

Parametrization

- Factorization property associated with conditional independence graph

$$p(\boldsymbol{z}, \boldsymbol{y}) = \prod_c [p(z_0^c) p(y_0^c | z_0^c)] \prod_{t>0} \prod_c [p(y_t^c | \mathrm{pa}(y_t^c)) p(z_t^c | \mathrm{pa}(z_t^c))$$

where $\mathrm{pa}(x)$ refers to the parents of x.

- Induces a canonical parametrization for $p(\boldsymbol{Z}, \boldsymbol{Y})$ in the case of discrete observed variables.
- Can be extended to continuous observed variables using regression models.
- Focusing on discrete variables for the sake of concision, does not change computational complexity of marginalization.

Example of full coupling states/observations:

$$\mathbb{P}(Z_t^c = j | Z_{t-1}^c = i, \boldsymbol{Z}_{t-1}^{-c} = \boldsymbol{\nu}, \boldsymbol{Y}_{t-1} = \boldsymbol{y}_{t-1}) = a_{i,\boldsymbol{\nu},\boldsymbol{y}_{t-1},j};$$
$$\mathbb{P}(Y_t^c = y | Z_t^c = j, \boldsymbol{Z}_{t-1} = \boldsymbol{i}, \boldsymbol{Y}_{t-1} = \boldsymbol{y}_{t-1}) = p_{\theta_{\boldsymbol{i},j,\boldsymbol{y}_{t-1}}}(y).$$

Impact of couplings on complexity: EM algorithm in Multichain HMMs

Example of full coupling

The EM algorithm consists in iterated maximizations (in λ) of the function:

$$Q(\lambda, \lambda^{(m)}) = E_{\lambda^{(m)}}\left[\log p_\lambda(\boldsymbol{Z}, \boldsymbol{y}) | \boldsymbol{y}\right] = \sum_c \sum_j \mathbb{P}_{\lambda^{(m)}}(Z_0^c = j | \boldsymbol{y}) \log \pi_j$$

$$+ \sum_c \sum_j \mathbb{P}_{\lambda^{(m)}}(Z_0^c = j | \boldsymbol{y}) \log p_{\theta_j}(y_0^c)$$

$$+ \sum_{t=1}^{T} \sum_c \sum_i \sum_{\boldsymbol{\nu}} \mathbb{P}_{\lambda^{(m)}}(Z_{t-1}^c = i, \boldsymbol{Z}_{t-1}^{-c} = \boldsymbol{\nu}, Z_t^c = j | \boldsymbol{y}) \log a_{i, \boldsymbol{\nu}, \boldsymbol{y}_{t-1}, j}$$

$$+ \sum_{t=0}^{T} \sum_c \sum_{\boldsymbol{i}} \sum_j \mathbb{P}_{\lambda^{(m)}}(Z_t^c = j, \boldsymbol{Z}_{t-1} = \boldsymbol{i} | \boldsymbol{y}) \log p_{\theta_{\boldsymbol{i}, j, \boldsymbol{y}_{t-1}}}(y_t^c).$$

$\Rightarrow$ This requires to compute marginal probabilities $\mathbb{P}_{\lambda^{(m)}}(Z_0^c = j | \boldsymbol{y})$

Computational complexity, what is at stake?

Assuming common state space in multichain with domain size K

- Marginal probabilities $\mathbb{P}_{\lambda^{(m)}}(Z_0^c = j|\boldsymbol{y})$ can be computed in HMMs by the *Forward-Backward* algorithm (prevents time complexity $\mathcal{O}(K^T)$)
- In classical monochain HMMs, this can be done in time $\mathcal{O}(TK^2)$
- In multichain HMMs, a naive application of Forward-Backward to the collapsed process (with a single multivariate state of domain size K^C) leads to complexity $\mathcal{O}(TK^{2C})$.

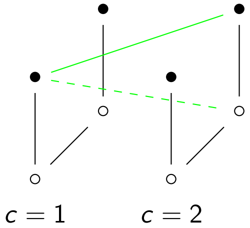
Thus, inference (and EM) is infeasible in this way when the number of hidden chains increases (event moderate C!)

$\Rightarrow$ Are there cases where this curse of dimentionality can be broken?

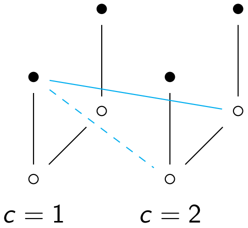
$\Rightarrow$ And when it cannot, are there efficiently computable approximations?

Reduced complexity due to conditional independence

Hidden variables from different chains may be independent given all observations $\boldsymbol{Y}$, when arcs linking nodes from different chains depart from observed variables only:



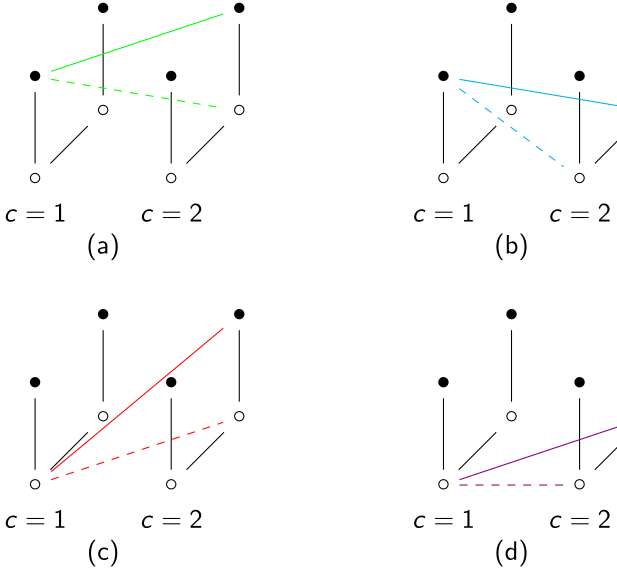
(a)



(b)

$\Rightarrow$ In this case, separate inference/EM algorithms can be applied to the chains, leading to overall complexity $\mathcal{O}(CTK^2)$.

Moralization graphs of the four types



Reduced complexity due to sparse transition matrices

- Forward/backward algorithm with C chains can be rewritten as a series of T matrix $\times$ vector products, with transition matrix $K^C \times K^C$, times vector of dimension K^C.
- Each product requires K^{2C} multiplications (why: as many multiplications as terms in the matrix), hence the global complexity in $\mathcal{O}(T\,K^{2C})$.
- The complexity of the matrix $\times$ vector product can be significantly decreased if the matrix is sparse: if the density of the matrix is in $K^{\rho C}$ with $\rho < 2$, the complexity is in $\mathcal{O}(T\,K^{\rho C})$.
- If $\rho < 1$, the graph of coupling is no longer connected, and there exists sets of independent chains (the number of edges is lower than K^C, the number of states); we give next an example of a model of metapopulation with $\rho = 1$.

Example of sparsity: a metapopulation model

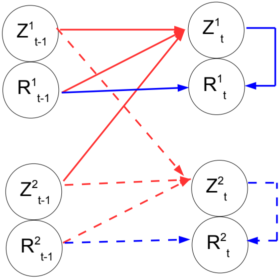
- Metapopulation: C patches, each patch is a chain with $K = 2$ states: occupied ($z = 1$) or empty ($z = 0$).
- at time t, transitions occur in one patch and one patch only selected randomly: colonisation if empty ($0 \to 1$) or extinction if occupied ($1 \to 0$).
- a simple example: $C = 3$, states $\mathbf{i}, \mathbf{j} \in \{0, 1\}^3$; transition matrix with $\times$ if $A[\mathbf{i}, \mathbf{j}] \neq 0$ and with "." if $A[\mathbf{i}, \mathbf{j}] = 0$; for example, $010 \to 011$ is a colonisation and $111 \to 000$ is impossible (3 events).

$\mathbf{i} \backslash \mathbf{j}$	000	001	010	111	100	101	110	111
000	$\times$	$\times$	$\times$	.	$\times$	.	.	.
001	$\times$	$\times$	.	$\times$	.	$\times$	.	.
010	$\times$	.	$\times$	$\times$	.	.	$\times$	.
011	.	$\times$	$\times$	$\times$	.	.	.	$\times$
100	$\times$	.	.	.	$\times$	$\times$	$\times$	.
101	.	$\times$	.	.	$\times$	$\times$	.	$\times$
110	.	.	$\times$	.	$\times$	.	$\times$	$\times$
111	.	.	.	$\times$	.	$\times$	$\times$	$\times$

One can show there are $\approx K^C$ non zero elements if $C \gg 1$.

Current work on multichain (hidden) semi-Markov processes and perspectives

- Core of the problem: changing sojourn duration underway.
- Current approaches restrict interactions at transition times Example of Touloupou *et al.* (2020), but ill-defined.
- Possibilities through discrete hazard rates $\lambda(d) = \mathbb{P}(D = d | D \geq d)$ ($= p$ in geometric $\mathcal{G}(p)$).
  - Firstly, introduce covariates $\lambda(d|x)$
  - Secondly, introduce states as covariates $\lambda_c(d|\boldsymbol{Z}_{t-1}^{-c})$ while guaranteeing finite D.



Example of redefinition of coupled SMMs introduced by Touloupou *et al.* (2020) using residual duration R_t; possible alternatives with elapsed times E_t.

Concluding remarks

- Multichain H(S)MMs: versatile framework for modelling various temporal processes on networks.
- Current and future work: catalogue of toolboxes, defining and learning couplings.
- Addressing computational complexity with approximate EM steps (variational EM, mean field) or Bayesian estimation.

References I

Brand, M., Oliver, N., & Pentland, A. (1997). Coupled hidden Markov models for complex action recognition. *Computer Vision and Pattern Recognition*, 994–999.

Cros, M.-J., Aubertot, J.-N., Peyrard, N., & Sabbadin, R. (2017). GMDPtoolbox: A Matlab library for designing spatial management policies. Application to the long-term collective management of an airborne disease. *PLOS One*, *12*(10), e0186014.

Ghahramani, Z., & Jordan, M. (1997). Factorial hidden Markov models. *Machine Learning*, *29*(2-3), 245–273.

Le Coz, S., Cheptou, P.-O., & Peyrard, N. (2019). A spatial Markovian framework for estimating regional and local dynamics of annual plants with dormancy. *Theoretical Population Biology*.

Touloupou, P., Finkenstädt, B., & Spencer, S. E. F. (2020). Scalable bayesian inference for coupled hidden markov and semi-markov models. *Journal of Computational and Graphical Statistics*, *29*(2), 238–249.

References II

Wainwright, M., & Jordan, M. (2008). Graphical models, exponential families, and variational inference. In *Foundations and trends in machine learning* (pp. 1–305).

Examples of applications on slide 8

(c) Phoma propagation; pathogenic fungi (Cros et al., 2017) is actually a MDP on a graph, with coupling of type (d)

(d) Coupled HMMs: spread of infection (Touloupou et al., 2020) is a true multichain HMM with coupling of type (d)