Hidden Markov Models

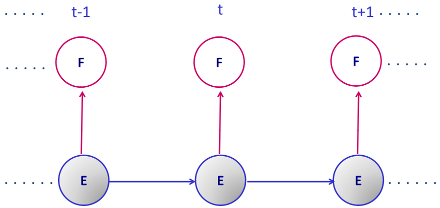
Sabrina Giordano

Dept of Economics, Statistics and Finance, University of Calabria - Italy

Kick-off Meeting for PRIN 2012

Roma, 12 Febbraio 2014

A hidden Markov model assumes the existence of an observed categorical time series $\mathbf{F} = \{F_t : t = 0, 1, 2...\}$ which depends on a latent process $\mathbf{E} = \{E_t : t = 0, 1, 2...\}$



- **E** is a first order Markov chain:

$$Pr(E_t|E_1, E_2, ..., E_{t-1}) = Pr(E_t|E_{t-1}) = \phi(e|e')$$

- local independence:

$$Pr(F_t|F_1, F_2, ..., F_{t-1}, E_1, E_2, ..., E_t) = Pr(F_t|E_t) = \varphi(f|e)$$

Latent and observation models

Latent model

- transition probability: $\phi(e|e')$

latent state: $e \in \mathcal{E}, e' \in \mathcal{E}$

Observation model

- state-dependent probability: $\varphi(f|e)$

observed values: $f \in \mathcal{F}$, latent state: $e \in \mathcal{E}$ and

Multivariate HMM

– An observed multivariate categorical time series

$$\mathbf{F}_{\mathcal{V}} = \{F_j(t) : t \in \mathbb{N}, j \in \mathcal{V}\},$$

– A latent multivariate categorical process

$$\mathbf{E}_{\mathcal{U}} = \{E_i(t) : t \in \mathbb{N}, i \in \mathcal{U}\},$$

◇ $\mathcal{V} = \{1, ..., s\}, \mathcal{U} = \{1, ..., r\}, \mathbb{N} = \{0, 1, 2, ..., \}$

◇ $E_i(t)$, $F_j(t)$ take values in sets $\mathcal{E}_i$, $\mathcal{F}_j$, $i \in \mathcal{U}$, $j \in \mathcal{V}$

◇ marginal processes $\mathbf{E}_{\mathcal{T}} = \{E_i(t) : i \in \mathcal{T}, t \in \mathbb{N}\}$, $\mathbf{F}_{\mathcal{R}} = \{F_j(t) : j \in \mathcal{R}, t \in \mathbb{N}\}$
$\mathcal{T} \subset \mathcal{U}$, $\mathcal{R} \subset \mathcal{V}$

Multivariate Hidden Markov Models

def. MHMM

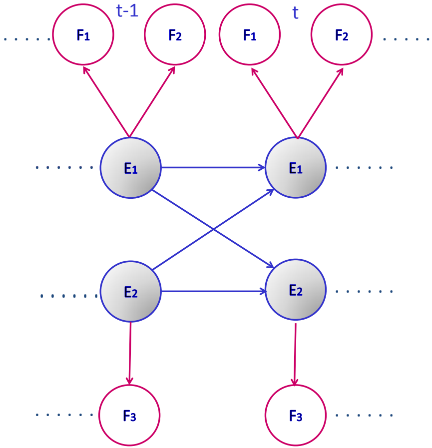
The joint process $(\mathbf{E}_{\mathcal{U}}, \mathbf{F}_{\mathcal{V}})$ is an MHMM if

- $\mathbf{E}_{\mathcal{U}}$ is a not observable Markov chain
- $F_{\mathcal{V}}(t) \perp\!\!\!\perp E_{\mathcal{U}}(t-1), F_{\mathcal{V}}(t-1) | E_{\mathcal{U}}(t)$

MHMMs are well suited to applications where:

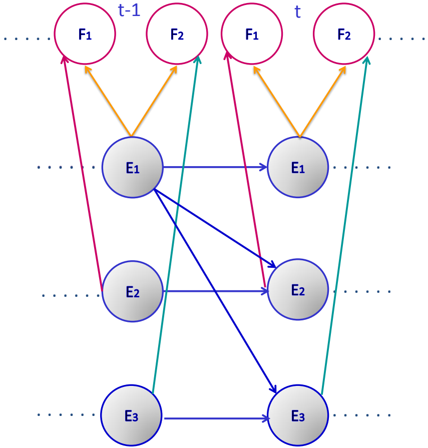
- the effect of a subset of latent variables is only on a subset of observed variables
- all the observed time series are affected by one common unobservable factor (general effect) and each observable variable is also governed by its specific latent variable
- an unobservable aspect influences the marginal dynamics of each observed variable while another latent factor influences the association among them
- one or more latent variables depend on the past of other latent variables (causality)
- contemporaneous association/independence among variables

Case 1



- Granger causality among latent var's; contemp. independence
- every observable var depends on one latent var, local independence

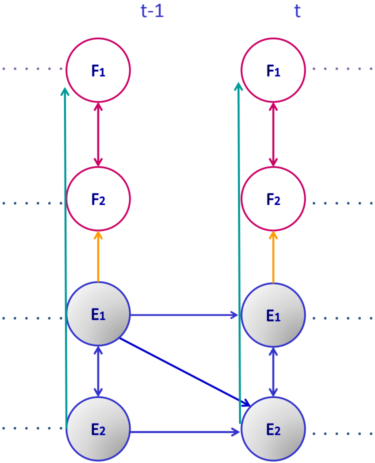
Case 2



◇ no G-causality among some latent var's; contemp. independence

◇ **generic effect**: one latent var. affects all the observable var's
specific effect: each observable var has its specific latent var; local ind.

Case 3



- ◇ each obs var. depends on one latent var., association among obs var's
- ◇ no G-causality of $\mathbf{E}_1$ by $\mathbf{E}_2$, association among latent var's

Latent and observation models

Latent model

- transition probability: $\phi(\boldsymbol{e}|\boldsymbol{e}')$

latent state: $\boldsymbol{e} = (e_1, e_2, ..., e_r)$, $\boldsymbol{e} \in \mathcal{E}$, $\boldsymbol{e}' \in \mathcal{E}$

Observation model

- state-dependent probability: $\varphi(\boldsymbol{f}|\boldsymbol{e})$

observed values: $\boldsymbol{f} = (f_1, f_2, ..., f_s)$, $\boldsymbol{e} \in \mathcal{E}$ and $\boldsymbol{f} \in \mathcal{F}$

Marginal models for MHMMs

Key point:

- the hypotheses of independencies/relations among obs and latent var's of MHMMs are restrictions on $\phi(\boldsymbol{e}|\boldsymbol{e}')$ and $\varphi(\boldsymbol{f}|\boldsymbol{e})$
- marginal parameterizations to model transition $\phi(\boldsymbol{e}|\boldsymbol{e}')$ and state-dependent probabilities $\varphi(\boldsymbol{f}|\boldsymbol{e})$
- the hypotheses are tested as constraints on parameters of marginal models

We adopt a Gloneck-McCullagh multivariate logistic model whose parameters $\lambda^{\mathcal{P}}(\boldsymbol{e}_{\mathcal{P}}|\boldsymbol{e}')$, $\eta^{\mathcal{P}}(\boldsymbol{f}_{\mathcal{P}}|\boldsymbol{e})$ are log-linear interactions not defined on the joint distributions $\phi(\boldsymbol{e}|\boldsymbol{e}')$ and $\varphi(\boldsymbol{f}|\boldsymbol{e})$, but on the marginal distributions $\phi_{\mathcal{P}}(\boldsymbol{e}_{\mathcal{P}}|\boldsymbol{e}')$, $\varphi_{\mathcal{P}}(\boldsymbol{f}_{\mathcal{P}}|\boldsymbol{e})$ of the var's in the set $\mathcal{P}$ (e.g. logits, log-odds ratios)

- **for latent model:**

$$\lambda^{\mathcal{P}}(\boldsymbol{e}_{\mathcal{P}}|\boldsymbol{e}') = \sum_{\mathcal{K}\subseteq\mathcal{P}} (-1)^{|\mathcal{P}\setminus\mathcal{K}|} \log \phi_{\mathcal{P}}(\boldsymbol{e}_{\mathcal{K}}, \boldsymbol{e^*}_{\mathcal{P}\setminus\mathcal{K}}|\boldsymbol{e}')$$

$\mathcal{P} \subseteq \mathcal{U}, \mathcal{P} \neq \emptyset, \boldsymbol{e}_{\mathcal{P}} \in \times_{i\in\mathcal{P}}\mathcal{E}_i$

- **for obs model:**

$$\eta^{\mathcal{P}}(\boldsymbol{f}_{\mathcal{P}}|\boldsymbol{e}) = \sum_{\mathcal{K}\subseteq\mathcal{P}} (-1)^{|\mathcal{P}\setminus\mathcal{K}|} \log \varphi_{\mathcal{P}}(\boldsymbol{f}_{\mathcal{K}}, \boldsymbol{f^*}_{\mathcal{P}\setminus\mathcal{K}}|\boldsymbol{e})$$

$\mathcal{P} \subseteq \mathcal{V}, \mathcal{P} \neq \emptyset, \boldsymbol{f}_{\mathcal{P}} \in \times_{j\in\mathcal{P}}\mathcal{F}_j$

Testing dynamic relations

G-noncausality and contemporaneous independence conditions in latent and observation models are equivalent to simple linear constraints on the $\lambda^{\mathcal{P}}$ and $\eta^{\mathcal{P}}$ parameters

For an MHMM with positive probabilities it holds that

- $E_{\mathcal{T}}(t) \perp\!\!\!\perp E_{\mathcal{U}\setminus\mathcal{T}}(t-1) | E_{\mathcal{T}}(t-1) \Leftrightarrow \lambda^{\mathcal{P}}(\boldsymbol{e}_{\mathcal{P}}|\boldsymbol{e}') = \lambda^{\mathcal{P}}(\boldsymbol{e}_{\mathcal{P}}|\boldsymbol{e}'_{\mathcal{T}})$

  $\mathcal{P} \subseteq \mathcal{T}, \mathcal{T} \subseteq \mathcal{U}$

- $F_{\mathcal{R}}(t) \perp\!\!\!\perp E_{\mathcal{U}\setminus\mathcal{T}}(t) | E_{\mathcal{T}}(t) \quad \Leftrightarrow \quad \eta^{\mathcal{P}}(\boldsymbol{f}_{\mathcal{P}}|\boldsymbol{e}) = \eta^{\mathcal{P}}(\boldsymbol{f}_{\mathcal{P}}|\boldsymbol{e}_{\mathcal{T}})$

  $\mathcal{P} \subseteq \mathcal{R}, \mathcal{T} \subseteq \mathcal{U}$

$\boldsymbol{f}_{\mathcal{P}} \in \times_{j\in\mathcal{P}}\mathcal{F}_j$, $\boldsymbol{e}_{\mathcal{P}} \in \times_{i\in\mathcal{P}}\mathcal{E}_i$, $\boldsymbol{e}_{\mathcal{T}} \in \times_{i\in\mathcal{T}}\mathcal{E}_i$

model estimations (EM algorithm) and tests by the *R*-package **hmmm** (Colombi et al, 2012)

Available from HMM

- conditional distributions $Pr(F_t|\mathbf{F}^{(-t)})$
- forecast distributions $Pr(F_{t+h}|\mathbf{F}^{(T)})$
- local decoding $Pr(E_t|\mathbf{F}^{(T)})$
- global decoding: max over $\mathbf{E}^{(T)}$ the prob $Pr(\mathbf{E}^{(T)}|\mathbf{F}^{(T)})$
- state prediction $Pr(E_{t+h}|\mathbf{F}^{(T)})$

THANKS FOR

YOUR ATTENTION!!

Our works

- COLOMBI R, GIORDANO S (2011). Testing lumpability for marginal discrete hidden Markov models. ASTA ADVANCES IN STATISTICAL ANALYSIS, vol. 95, p. 293-311
- COLOMBI R, GIORDANO S (2013). Multiple hidden Markov models for categorical time series. Submitted.
- COLOMBI R, GIORDANO S, CAZZARO M (2012). R-package hmmm: hierarchical multinomial marginal models. In: CRAN R-package

EXAMPLE:
a one-year time series of daily sales of soft-drinks:
lemon tea $\mathbf{F}_T$, orange and apple juices $\mathbf{F}_O, \mathbf{F}_A$ (low, medium, high level);
2-states $\mathbf{E}_1, \mathbf{E}_2$

MHMM for soft-drinks data

◇ **latent model hypotheses**

- $\mathbf{E}_1$ and $\mathbf{E}_2$ are marginally Markov chains:

$$E_1(t) \perp\!\!\!\perp E_2(t-1)|E_1(t-1), \quad E_2(t) \perp\!\!\!\perp E_1(t-1)|E_2(t-1)$$

- constant association: $\lambda^{\mathcal{P}}(\boldsymbol{e}_{\mathcal{P}}|\boldsymbol{e}') = \lambda^{\mathcal{P}}(\boldsymbol{e}_{\mathcal{P}})$ for $|\mathcal{P}| > 1$

◇ **observation model hypotheses**

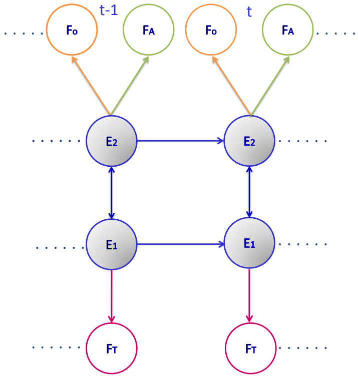
- $\mathbf{F}_T$ depends on $\mathbf{E}_1$; and $(\mathbf{F}_O, \mathbf{F}_A)$ on $\mathbf{E}_2$:

$$F_T(t) \perp\!\!\!\perp E_2(t)|E_1(t), \quad F_O(t), F_A(t) \perp\!\!\!\perp E_1(t)|E_2(t)$$

- local independence:

$$F_T(t) \perp\!\!\!\perp F_O(t) \perp\!\!\!\perp F_A(t)|E_1(t), E_2(t)$$

MHMM for soft-drinks data



- test results $G^2 = 94.1538$, $df = 99$, $p\text{-}value = 0.6189$
- data, model estimations (EM algorithm) and tests by the *R*-package **hmmm** (Colombi et al, 2012)

in the univariate case:

- $Pr(F) = \sum_{\mathcal{E}} P(F|E)Pr(E)$
- $Pr(F_1, F_2, ..., F_T|E_1, E_2, ..., E_T) = \prod_{t=1}^{T} P(F_t|E_t)$
- $Pr(E_1, E_2, ..., E_T) = P(E_1)\prod_{t=2}^{T} P(E_t|E_{t-1})$
- $Pr(F_1, F_2, ..., F_T) = \sum_{\mathcal{E}} P(E_1)\prod_{t=2}^{T} P(E_t|E_{t-1})\prod_{t=1}^{T} P(F_t|E_t)$