Review of estimation algorithms of HMM/HSMM with mixed effects

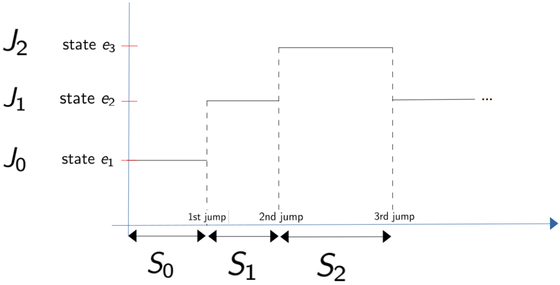
Mathieu VALDEYRON (PhD Student)

MIAT, INRAE, Toulouse, France

3 July 2025

Recap HSMM (explicit duration)

S_n = **dwell time** of the n-jump
J_n = the **n-state** visited



Markov chain : $P(S_n|J_n = k) =$ Geometric $(p_{\mathbf{k}})$

Semi-Markov chain (explicit duration) : $P(S_n|J_n = k) =$ any law, Poisson $(\lambda_{\mathbf{k}})$ +1 for example

Variability impact of a qualitative covariate (as in GLMM)

Qualitative covariate's effects can be traited as random, when we are not particularly interested in the effect of each modality, but rather in the law of the effects.

Taking account long-range dependency (in HMM) [2]

The observed variables $(Y_t)_t$ are already not independent without random effects, but

$$\mathrm{Cov}(Y_t, Y_{t+h}) \underset{h \to +\infty}{\longrightarrow} 0$$

And with random effects in the model, under some (restrictive) conditions

$$\mathrm{Cov}(Y_t, Y_{t+h}) \underset{h \to +\infty}{\longrightarrow} k > 0$$

Notation : U = vector of all random effects.
For individual/trajectory number i, Y^i = **observed process**,
Z^i = **hidden process**.

Definition

- U is a centred Gaussian vector (most of the time)
- Given $U = u$ and the covariates,
  1. $\{(Y^i, Z^i)\}_{1 \leq i \leq \mathcal{I}}$ are **mutually independent**
  2. (Y^i, Z^i) is an **H(S)MM** impacted by u and the covariates of individual number i
  3. Every law that compose the H(S)MM (for explicit duration case) :
     - emissions laws
     - transitions laws
     - dwell time laws
     - initial state law

     belong to **the exponential family**, and the expected value **depends linearly** on u and covariates through a link function.

Example of HMM, with random effect specific to individual and state

- **Emissions laws :**

$$f(y_t^i|Z_t^i = k, U = u) = \exp\left(\frac{y_t^i \eta_{k,u}^i - b(\eta_{k,u}^i)}{\phi_k} - c(y_t^i, \phi_k)\right)$$

$$\eta_{k,u}^i = x^{i\prime}\beta_k^{(1)} + u_k^{i,(1)} \qquad \text{with } u_k^{i,(1)} \sim \mathcal{N}(0, \sigma^2_{k,(1)})$$

$\longrightarrow$ In each state k, we have a different GLMM regression.

- **Hidden process laws:**

$$P(Z_{t+1}^i = \mathbf{l}|Z_t^i = k, U = u) = \frac{\exp\left(x^{i\prime}\beta_{k,\mathbf{l}}^{(2)} + u_{k,\mathbf{l}}^{i,(2)}\right)}{1 + \sum_\alpha x^{i\prime}\beta_{k,\alpha}^{(2)} + u_{k,\alpha}^{i,(2)}}$$

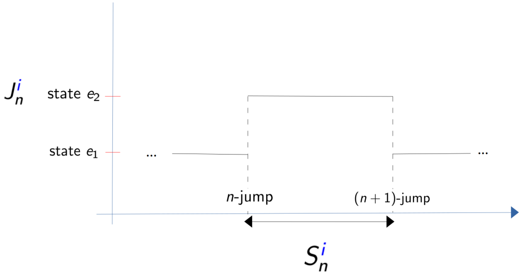
$$\text{with } u_{k,\mathbf{l}}^{i,(2)} \sim \mathcal{N}(0, \sigma^2_{k,\mathbf{l},(2)})$$

$\longrightarrow$ In each state k, we have a different GLMM (multinomial) regression.

Each article is modelled as a special case, or with a design vector for random effects, or with dynamical covariates.

Example of HSMM, for dwell time only with random effect specific to individual and state

For the individual/trajectory number i :
S_n^i = **dwell time** of the n-jump of the hidden process
J_n^i = the **n-state** visited by the hidden process.



- **dwell time, modelled by a Poisson:**

$$P(S_n^i | J_n^i = k, U = u) = \mathcal{P}(\lambda_{k,u}^i) + 1$$

$$\lambda_{k,u}^i = \exp\left(x^{i\prime}\beta_k + u_k^i\right) \qquad \text{with } u_k^i \sim \mathcal{N}(0, \sigma_k^2)$$

Estimation problems

- The integration over random effects is intractable.
- Fixed effects also introduce maximisation difficulties.

These are problems derived from GLMM.

Several estimation methods are available.

EM	Newton-Raphson on the likelihood	Likelihood approximation, and maximisation	Bayesian
- **Altman** (2007) Monte-Carlo EM (MCEM) **HMM** - **Chaubert** (2008) Restoration-Maximisation, with Monte-Carlo **HSMM** - **Maruotti** (2011) no detail **HMM** - **Delattre, Lavielle** (2012) Stochastique Approximation EM (SAEM) (Monte-Carlo) **HMM**	**Haji-Maghsoudi & al.** (2021) Monte-Carlo Newton-Raphson **HSMM**	- **Altman** (2007) Gaussian quadrature **HMM** - **Altman** (2007) Importance sampling **HMM** - **Michelot** (2023) Laplace quadrature **HMM** - **Koslik** (2024) Laplace quadrature with quasi-REML in addition (integration over fixed effects of the likelihood) **HMM**	- **Michelot** (2023) MCMC Hamiltonian **HMM**

Notation : Y = **all observed process**, Z = **all hidden process**, U = **all random effects**, f^θ = **density function**.

EM

step E

$$Q(\theta|\theta^{(m)}) = E_{\theta^{(m)}}\left[\log f^\theta(y, Z, U)|Y = y\right]$$
$$= E_{\theta^{(m)}}\left[\underbrace{E_{\theta^{(m)}}\left[\log f^\theta(y, Z|U)|U, Y = y\right]}_{=:h_{\theta^{(m)}}(\theta,U,y)}|Y = y\right] + E_{\theta^{(m)}}\left[\log f^\theta(U)|Y = y\right] \tag{1}$$

Integration over Z is possible : the function $h_{\theta^{(m)}}(\theta, u, y)$ can be calculated analytically by forward-backward for a given u.

$\longrightarrow$ Need to approximate integration over U.

step M

Maximisation by Newton-Raphson (or quasi-Newton) most of the time

EM	**Altman** (2007)	**Chaubert** (2008)	**Maruotti** (2011)	**Delattre, Lavielle** (2012)
	HMM	HSMM	HMM	HMM
random effetcs on **observed** process	*V*	*V* (gaussian emmision only)	*V*	*V*
random effetcs on **hidden** process	*V*	*X*	*X*	*V*
Dynamical covariates	*V*	*X*	*V*	*X*
E step	Monte-Carlo EM (MCEM)	Restoration-Maximisation, with Monte-Carlo at times	- Gaussian random effects : no details - Finite random effects : leads to special mixture of HMM.	Stochastic Approximation EM (SAEM) (Monte-Carlo). Convex combination between $\hat{Q}(.\|\theta^{(m-1)})$ and new simulation under $\theta^{(m)}$
M step	quasi-Newton	closed form (because gaussian emissions)	no details	no details

Notation : $Y =$ **all observed process**, $Z =$ **all hidden process**, $U =$ **all random effects**, $f^\theta =$ **density function**.

Newton-Rahpson

$$\nabla_\theta \log f^\theta(y) = E_\theta \Big(\nabla_\theta \log f^\theta(y, Z, U) \mid Y = y \Big)$$

$$\nabla^2_\theta \log f^\theta(y) = E_\theta \Big(\nabla^2_\theta \log f^\theta(y, Z, U) \mid Y = y \Big) + \mathrm{Var}_\theta \Big(\nabla_\theta \log f^\theta(y, Z, U) \mid Y = y \Big)$$

Like EM, integration over Z is analytically possible with forward-backward, with the exception of the variance term in the Hessian (in fact I don't know)

$\longrightarrow$ Need to approximate integration over U.

Newton-Rahpson

Haji-Maghsoudi & **al.** (2021)
Monte-Carlo Newton-Raphson
HSMM, random effects on observed process.

Notation : Y = **all observed process**, Z = **all hidden process**, U = **all random effects**, f^θ = **density function**.

Likelihood approximation, and maximisation

$L(\theta, y)$ is an approximation of the likehood $f^\theta(y) = \int f^\theta(y, u) \mathrm{du}$

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\, L(\theta, y)$$

$f^\theta(y, u) = f^\theta(y|u) f^\theta(u)$ with $f^\theta(y|u)$ calculable by forward.

Maximisation Newton-Raphson (or quasi-Newton) most of the time.

Likelihood approximation, and maximisation

	Altman (2007)	**Altman** (2007)	**Michelot** (2023)	**Koslik** (2024)
	HMM	HMM	HMM	HMM
random effetcs on **observed** process	V	V	V	V
random effetcs on **hidden** process	V	V	V	V
Dynamical covariates	V	V	V	V
Approximation's type	Gaussian quadrature + quasi-Newton	Importance sampling on $\int f^{\theta}(y, u)\mathrm{d}u$ + quasi-Newton	Laplace quadrature on $\int f^{\theta}(y, u)\mathrm{d}u$ + package optimix for maximisation	Laplace quadrature on $\int f^{\theta}(y, u)\mathrm{d}u\mathrm{d}\beta$ (with $\theta = (\beta, \sigma^2)$, β = fixed effects)

Things to remember

- HSMMs are not frequently consider
- EM and likelihood approximation well represented, unlike Newton-Raphson and Bayesian
- No realy comparison between methods, except :
  - few random effects per individual => small integration dimension : **quadrature** are efficient (faster than Monte-Carlo)
  - many random effects per individual => large integration dimension: **Monte-Carlo** to be preferred
- To eliminate the hidden process we can perform **forward-backward** (when taking the expectation in EM), or simply **forward** (when integrating in likelihood approximation).

Perspectives (of my PhD thesis)

1. consider HSMMs (with a general model)
2. implement some available methods (Monte-Carlo, quadrature, variationnal EM, Newton-Raphson, etc.)
3. compare estimation methods

[1] Haji-Maghsoudi et al. "Generalized linear mixed hidden semi-Markov models in longitudinal set-tings : A Bayesian approach". In: (2021).

[2] Altman. "Mixed hidden Markov models: An extension of the hidden Markov model to the longitudinal data setting". In: *Journal of the American Statistical Association* 102 (2007), pp. 201–210.

[3] Chaubert. "Combinaisons markoviennes et semi-markoviennes de mod'eles de r egression.Application 'a la croissance d'arbres forestier". PhD thesis. 2008.

[4] Lavielle Delattre. "Maximum likelihood estimation in discrete mixed hidden Markov modelsusing the SAEM algorithm". In: (2012).

[5] Koslik. "Efficient smoothness selection for nonparametric Markov-switching mo-dels via quasi restricted maximum likelihood". In: (2024).

[6] Maruotti. "Mixed Hidden Markov Models for Longitudinal Data : An Overview". In: (2011).

[7] Michelot. "hmmTMB : Hidden Markov models with flexible covariate effects in". In: (2023).