PhD Research Update

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

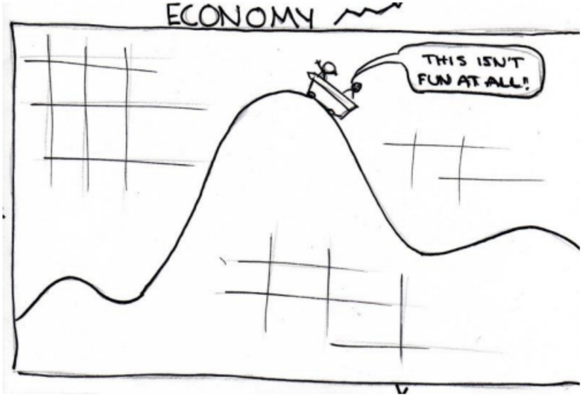
Market Changes, Moving Cycles
Bayesian Inference for hidden semi-Markov models

Patrick Aschermayr
Supervisor: Dr. Kostas Kalogeropoulos
Co-Supervisor: Prof. Pauline Barrieu

May 2019

Department of Statistics

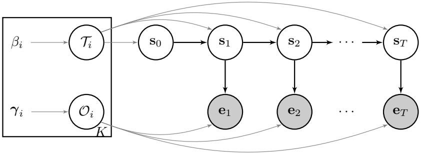
Motivation - foreword



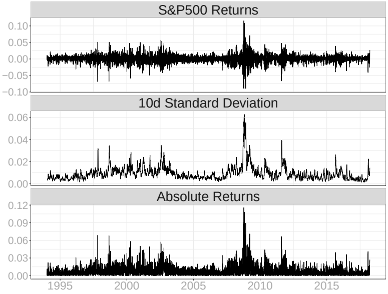
Motivation - basic idea

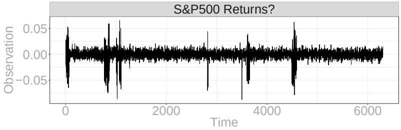
Definition: Hidden Markov model

A hidden Markov model(HMM) is a bivariate stochastic process $\{E_t, S_t\}_{t=1,2,\ldots}$, where $\{S_t\}$ is an unobserved Markov chain and, conditional on $\{S_t\}$, $\{E_t\}$ is an observed sequence of independent random variables.



K-state Bayesian HMM, parameter $\theta = \{\mathcal{T}, \mathcal{O}\}$ and hyperparameter $\{\beta, \gamma\}$.





(TOP): real data, (BOTTOM): HMM output

(Rabiner, 1989)

Table of contents

Challenges and research focus

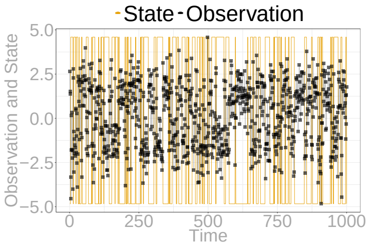
Basic HMM weaknesses

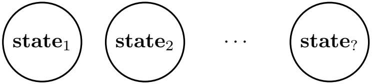
(1) **the implicit state geometric distribution** causes rapid switching between states.

→ explicitly model state durations, such models are known as hidden semi-Markov models (HSMM).

(2) the **a priori assignment of a fixed number of hidden states**.

→ Bayesian (nonparametric) framework (see Beal et al., 2002; Teh et al., 2006), for inferring arbitrarily large state complexity from data.



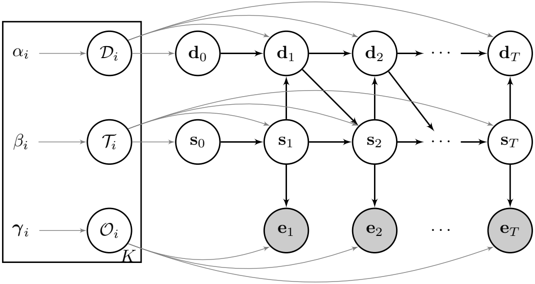


Bayesian hidden semi-Markov model

Model Definition

Definition: Hidden semi-Markov model

A hidden semi-Markov model is a bivariate stochastic process $\{Z_t, E_t\}_{t=1,2,\ldots,T}$, where $\mathbf{Z} = \{S, D\} = \{S_t, D_t\}_{t=1,2,\ldots,T}$ is an unobserved semi-Markov chain and, conditional on $\{Z_t\}$, $\{E_t\}$ is an observed sequence of independent random variables. S is the latent state sequence, and D is the corresponding 'latent remaining' duration sequence.



K-state Bayesian HSMM, parameter $\theta = \{\mathcal{D}, \mathcal{T}, \mathcal{O}\}$ and hyperparameters $\{\alpha, \beta, \gamma\}$.

Bayesian parameter estimation sketch

Target full posterior: $P(Z, \theta \mid E)$ **by iterating:**

1. Propose $\theta^\star$ from some proposal distribution $f(\theta^\star \mid \theta)$
2. Propose $Z^\star$ from the conditional distribution $P(Z^\star \mid \theta^\star, E)$.
3. Accept the pair $(\theta^\star, Z^\star)$ with acceptance probability

$$\begin{aligned} Acceptance &= \frac{P(Z^\star \mid \theta^\star)}{P(Z \mid \theta)} \frac{P(E \mid Z^\star, \theta^\star)}{P(E \mid Z, \theta)} \frac{P(Z \mid E, \theta)}{P(Z^\star \mid E, \theta^\star)} \frac{P(\theta^\star)}{P(\theta)} \frac{q(\theta \mid \theta^\star)}{q(\theta^\star \mid \theta)} \\ &= \frac{P(E \mid \theta^\star)}{P(E \mid \theta)} \frac{P(\theta^\star)}{P(\theta)} \frac{q(\theta \mid \theta^\star)}{q(\theta^\star \mid \theta)}, \end{aligned}$$

$$\begin{aligned} P(E \mid \theta) &= \sum_D \sum_S P(S, D, E \mid \theta) \\ &= \sum_D \sum_S P(s_0 \mid \mathcal{T}_0) P(d_0 \mid \mathcal{D}_0) \prod_{t=1}^{T} P(s_t \mid s_{t-1}, d_{t-1}, \mathcal{T}) P(d_t \mid s_t, d_{t-1}, \mathcal{D}) P(e_t \mid s_t, \mathcal{O}) \end{aligned}$$

→ Exact evaluation: computational complexity of up to $\mathcal{O}(K^2 T^3)$.
→ replace likelihood $\mathcal{L}_\theta(e_{1:T})$ with unbiased $\hat{\mathcal{L}}_\theta(e_{1:T})$ (Andrieu and Roberts, 2009)

Particle MCMC

Particle MCMC Algorithm

input : Proposal distribution Q, iterationNumber N,
Particle filter proposal π, particleNumber M,
observation $e_{1:T}$

output: $(\theta^i, Z^i_{1:T})_{i=1:N}$

Initialize θ ;

Run particle filter $\rightarrow$ *get* $\hat{P}(e_{1:T} \mid \theta)$*.* ;

for $i \leftarrow 1$ **to** N **do**

1. Propose a new $\theta^\star$, $\theta^\star \sim Q(\theta^\star \mid \theta)$;

2. Run particle filter $\rightarrow$ *get* $\hat{P}(e_{1:T} \mid \theta^\star)$ *and* $Z^\star_{1:T}$*.* ;

3. Accept the pair $(\theta^\star, Z^\star_{1:T})$ *with probability:*

$$min(1, \frac{\hat{P}(e_{1:T} \mid \theta^\star)}{\hat{P}(e_{1:T} \mid \theta)} \frac{P(\theta^\star)}{P(\theta)} \frac{Q(\theta \mid \theta^\star)}{Q(\theta^\star \mid \theta)})$$

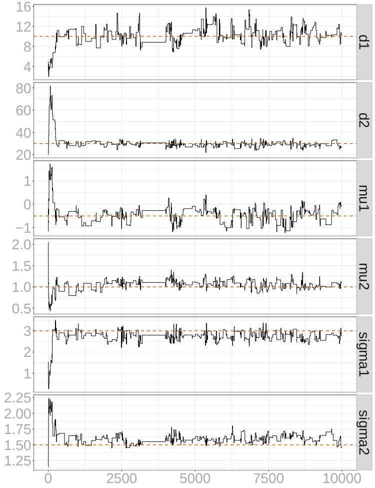
4. If accepted, set $\hat{P}(e_{1:T} \mid \theta) = \hat{P}(e_{1:T} \mid \theta^\star)$ *and* $\theta = \theta^\star$*.*

end

Algorithm 1: Particle Metropolis Hastings algorithm

Particle MCMC illustration

Posterior samples:



Filtering estimates:

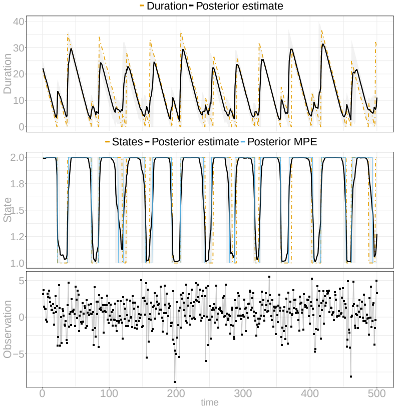


Fig.1 Posterior mean remaining state duration with 10% & 90% credible intervals against actual duration.
Fig.2 Posterior mean and most probable visited state against actual state.
Fig.3 Observation sequence input for particle MCMC sampler.

Particle Filter - likelihood approximation

Filtering sketch - Approximate C_t, $\hat{C}_t = \frac{1}{N}\sum_{n=1}^{N}\prod_{t=1}^{T} w_t^n(X_{1:t}^n)$

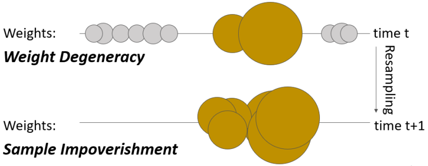
1. Goal: Sample from $P_t(X_{1:t}) = \frac{\tau_t(X_{1:t})}{C_t}$.
2. Choose $\pi_t(X_{1:t}) \propto \tau_t(X_{1:t})$ of form $\pi_t(X_{1:t}) = \pi_{t-1}(x_{1:t-1})\pi_t(X_t)$.
3. Iterate for $k = 1, ..., t$:
Sample (i) I particles $X_k^i \sim \pi_k(X_k \mid x_{1:k-1}^i)$.
Get $w_t^i(X_{1:t}^i) = w_{t-1}^i(X_{1:t-1}^i)\frac{\tau_t(X_t^i|x_{1:t-1}^i)}{\pi_t(X_t^i|x_{1:t-1}^i)}$, $W_t^i = \frac{w_t^i(X_{1:t}^i)}{\sum_n w_t^n(X_{1:t}^n)}$.
Resample & reweight $X_{1:k}^i$ according to W_t^i.

HSMM: $P_t(X_{1:t}) = P(Z_{1:t} \mid e_{1:t})$

$$P_t(Z_{1:t} \mid e_{1:t}) = \frac{\prod_{t=1}^{T} P(Z_t|Z_{t-1})P(e_t|Z_t)}{P(e_{1:T})}$$
$$w_t(X_{t-1:t}) = \frac{P(Z_t|Z_{t-1})P(e_t|Z_t)}{P(Z_t|e_t,Z_{t-1})}$$

$\rightarrow$ (ii) $\pi_t \approx P(Z_t \mid Z_{t-1}, e_t)$ unfeasable
$\rightarrow$ select $\pi_t \approx P(Z_t \mid Z_{t-1})$

(iii) Resampling trade-off:

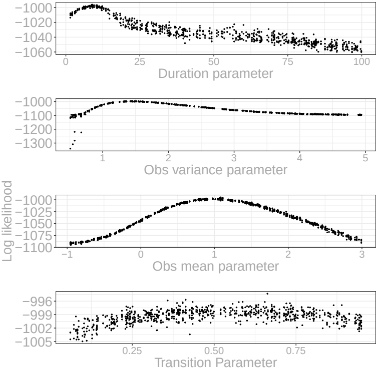


Particle Filter - Proposal distribution verification

Questions:

- Likelihood function peaked?
- $\mathrm{Var}[\hat{p}_\theta(e_{1:T})]$ constant across θ?

HSMM likelihood estimate on a θ grid:



Conclusion:

- Likelihood function moderately peaked.
- $\mathrm{Var}[\hat{p}_\theta(e_{1:T})]$ not independent of $\mathcal{D}$.

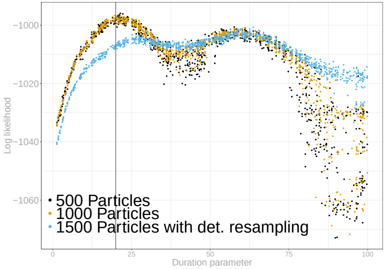
$\rightarrow$ Increase number of particles.

$\rightarrow$ Tune particle size to $\mathrm{Var}[\hat{p}_\mathcal{D}(e_{1:T})]$.

$\rightarrow$ Find better $\pi_t \approx P(Z_t \mid Z_{t-1}, e_t)$.

$\rightarrow$ ”Jitter” particles.

HSMM likelihood estimate for duration grid:



Preliminary results and outlook

Bayesian inference for HSMMs - status quo

Advantage particle MCMC:

1. Computational complexity both linear in time T and in number of particles N, $\mathcal{O}(NT)$, $N \approx T$.

2. Computational complexity independent of number of states.

3. Straight forward to extend estimation to the continuous state case.

Challenges:

1. Particle Filter:
$\rightarrow$ Find better $\pi_t \approx P(Z_t \mid Z_{t-1}, e_t)$.
$\rightarrow$ "Jittering" particles.

2. PMCMC Sampler:
$\rightarrow$ Find better $Q(\theta^\star \mid \theta)$.
$\rightarrow$ Initialize θ.

3. Infinite HSMMs

4. (Sequential parameter estimation framework)

Discussion

Appendix

References

Andrieu, C. and Roberts, G. O. (2009). The pseudo-marginal approach for efficient monte carlo computations. *Ann. Statist.*, 37(2):697–725.

Beal, M., Ghahramani, Z., and Rasmussen, C. (2002). The infinite hidden markov model. *Advances in neural information processing systems*, 14:577–584.

Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286.

References II

Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. (2006). Hierarchical dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581.