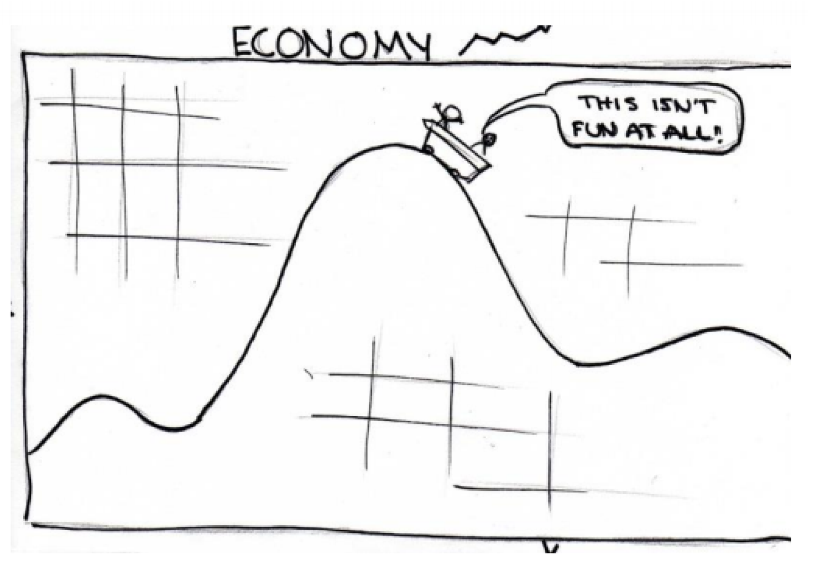
Market Changes, Moving Cycles - Bayesian inference for hidden semi-Markov models

Patrick Aschermayr
p.aschermayr@lse.ac.uk
Department of Statistics, London School of Economics

Motivation

Economies and financial markets move in cycles. A long period of expansion is often interrupted by an abrupt shock, followed by a prolonged recession period. Government policy, economic decision-making and several areas in financial institutions could all significantly benefit from advancement in this research area. How could one model such phenomena?

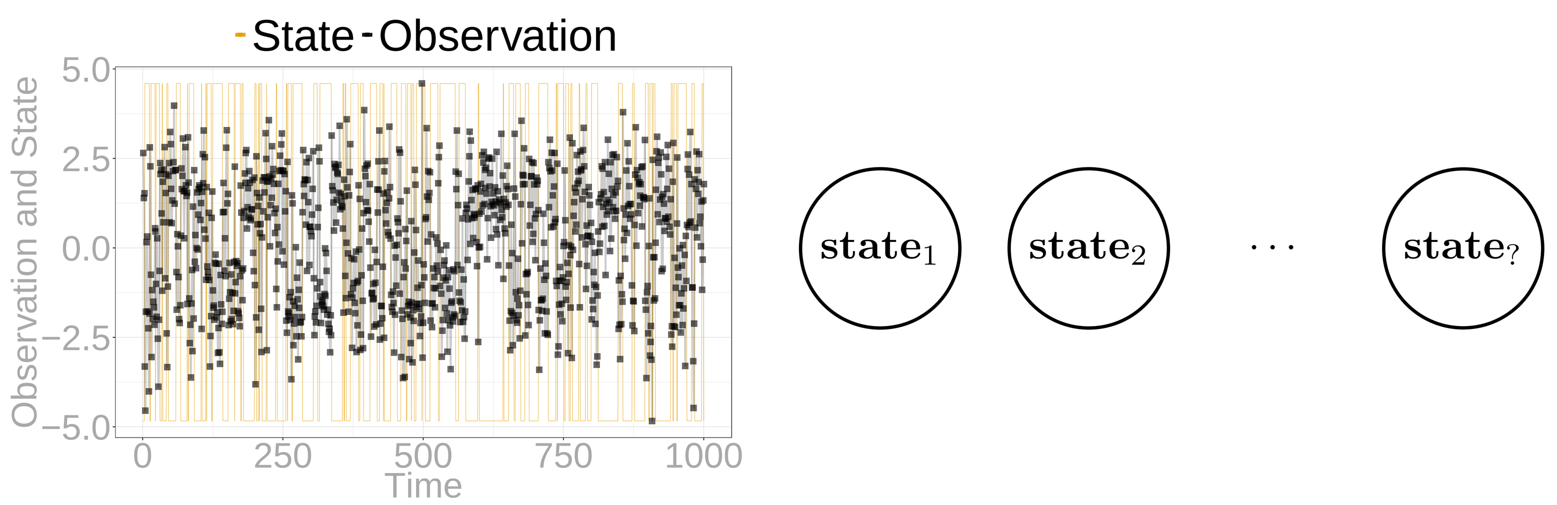


Status quo

As a starting point, basic hidden Markov models (HMM) have often been used to describe economic behavior. At each time step, a latent state variable is assigned to a (multivariate) observation sequence. Major weaknesses of such models are

(1) **the implicit geometric distribution** of HMMs, which causes rapid switching between states:

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



Problem 1 can be circumvented by explicitly modelling state durations, such models are known as hidden semi-Markov models (HSMM). Problem 2 may be tackled in a Bayesian (nonparametric) framework (see Beal et al., 2002; Teh et al., 2006), for inferring arbitrarily large state complexity from data. Both extensions come at the cost of higher computational burden and less tractability.

Challenges and research focus

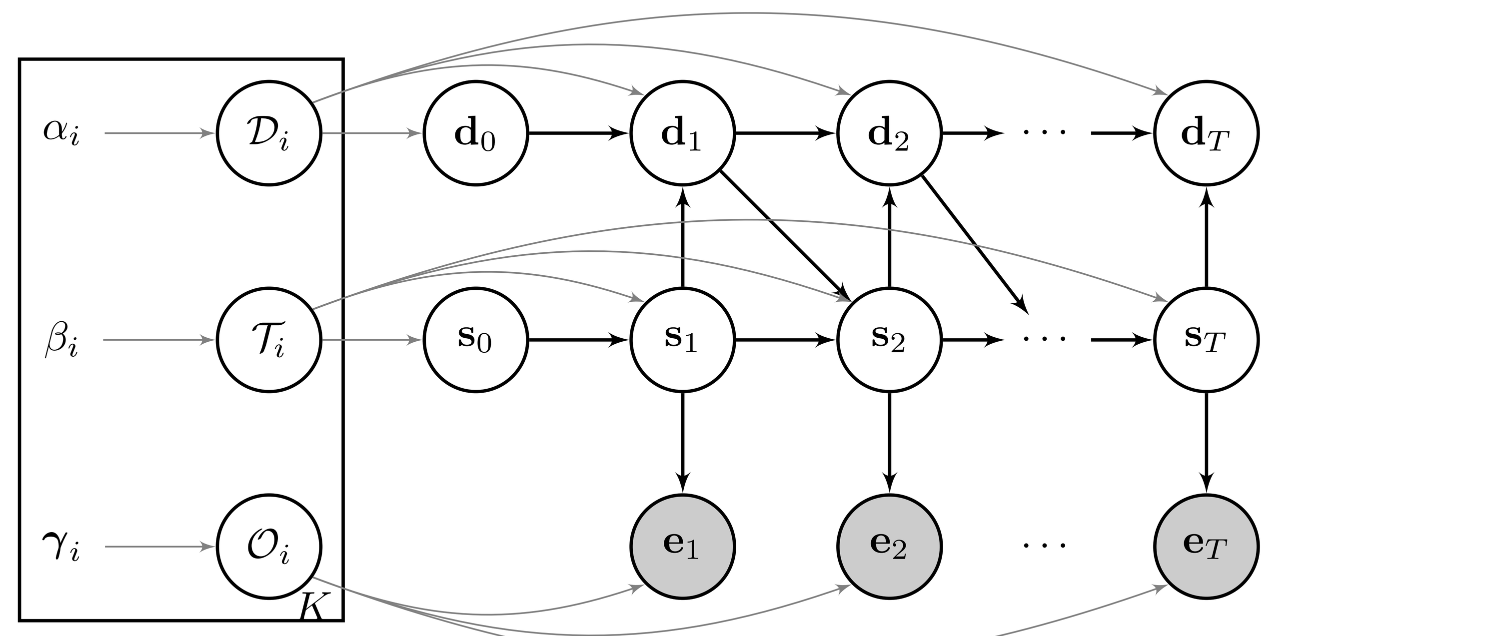
Unfortunately, current Bayesian implementations of HSMMs typically suffer from various difficulties:

- Metropolis style MCMC sampler need to evaluate likelihoods, which is computationally challenging for HSMMs. Iterative forward-backward algorithms have a time complexity of $\mathcal{O}(K^2(d_{max}-d_{min})^2T)$, where T is the sequence length, K the number of states and d the duration length. Typically, $(d_{max}-d_{min}) >> K$ can range from 1 up to T.
- Gibbs-style MCMC implementations typically suffer from a large autocorrelation in the posterior samples as parameter and state trajectories are highly correlated.

My research focus is thus on simultaneously tackling both weaknesses of basic HMMs by implementing efficient Bayesian inference algorithms for HSMMs.

Bayesian hidden semi-Markov model

Definition 1 (Hidden semi-Markov model). *A hidden semi-Markov model is a bivariate stochastic process $\{Z_t, E_t\}_{t=1,2,...,T}$, where $\mathbf{Z} = (S, D) = \{S_t, D_t\}_{t=1,2,...,T}$ 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 (α, β, γ).

Parameter estimation

In a Bayesian setting, one is interested in the full posterior $Z = (S, D)$ and θ given the observations E, $\mathcal{P}(Z, \theta \mid E)$. As no closed form solution exists, one typically resorts to MCMC algorithms to draw samples from this distribution. A standard MCMC sampler can be implemented by repeating N times the following steps:

1. Propose $\theta^\star$ from a 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

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

where one can simplify the acceptance probability using the basic marginal likelihood identity (BMI) of Chib (1995). The main challenge here is to evaluate the (marginal) likelihood $P(E \mid \theta)$,

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

as both S and D are unknown and standard iterative algorithm used in the basic HMM case have a time complexity of up to $\mathcal{O}(K^2T^3)$. An alternative would be to replace this (marginal) likelihood $\mathcal{L}_\theta(e_{1:T})$ with an unbiased estimate $\hat{\mathcal{L}}_\theta(e_{1:T})$. (Andrieu and Roberts, 2009) have shown the puzzling result that one can do so and still target the exact posterior distribution of interest, which led to a new research area now called 'exact approximate MCMC'. MCMC algorithm used in this setting are known as 'particle MCMC' sampler, see Kantas et al. (2014).

Particle MCMC

To summarize, a particle filter is integrated in a MCMC algorithm to obtain $\hat{\mathcal{L}}_{\theta^\star}(e_{1:T})$ and to draw a trajectory sample $Z^\star_{1:T}$. The computational complexity is both linear in time T and in number of particles N, $\mathcal{O}(NT)$, $N \approx T$. It is straight forward to extend estimation to the continuous state case and the computational complexity is not dependent on the number of states. A summary plot for posterior statistics of a particle metropolis sampler for a 2-state HSMM with normal observations and poisson durations is shown below.

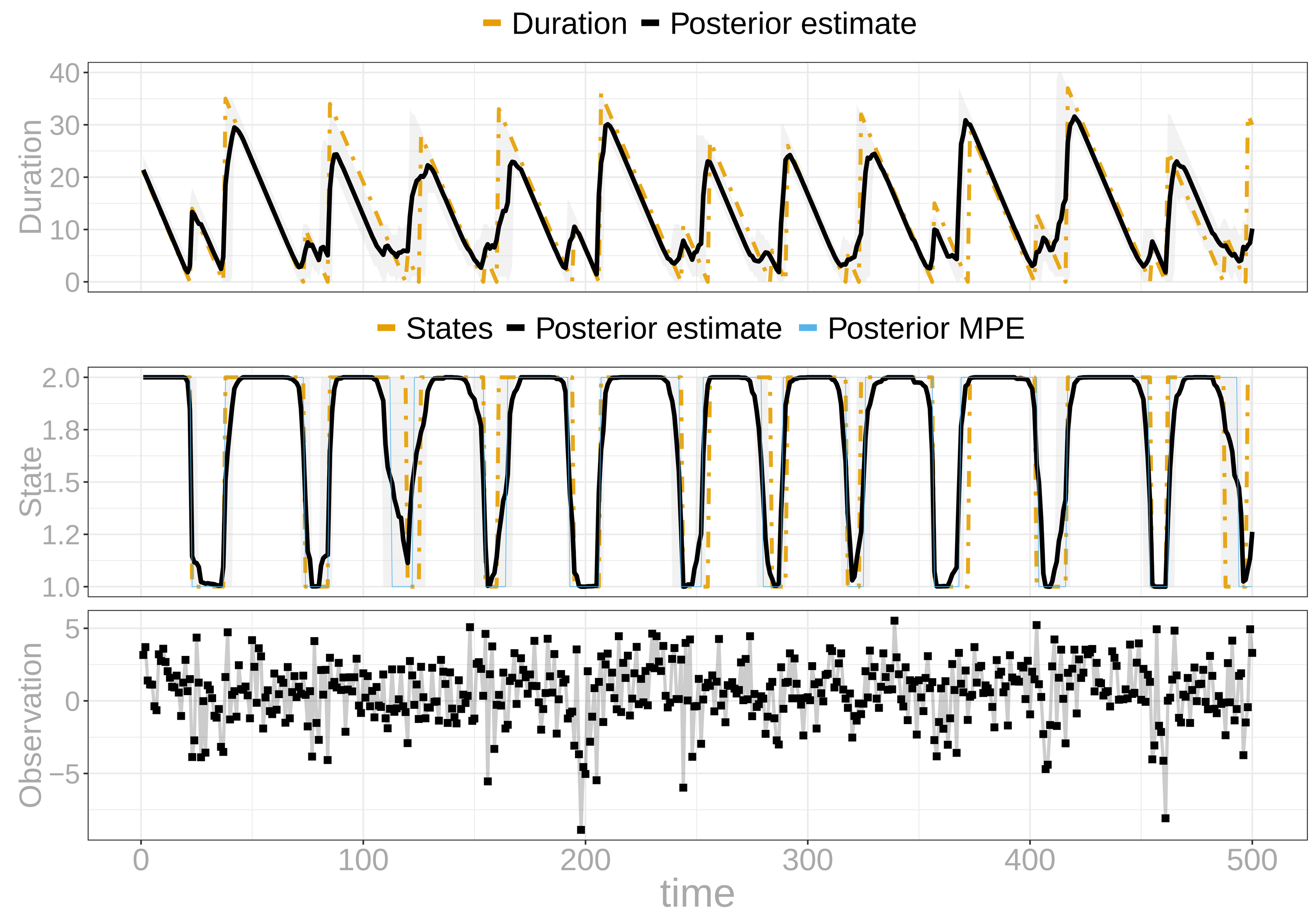


Fig.1 shows the average posterior remaining state duration and the 10% & 90% credible interval at each time step against the actual duration. Fig.2 shows the posterior mean and the most probable visited state against the actual state. Fig.3 shows the observation sequence that served as input for the particle MCMC sampler.

Preliminary results and outlook

- Particle MCMC is a competitive alternative to current Bayesian inference algorithm for HSMMs. Most notably, the time complexity of the likelihood evaluation is theoretically not dependent on the number of states, which makes particle MCMC suitable to simultaneously tackle both basic HMM problems discussed in the previous section.
- Practically, particle filter often experience path degeneracy problems. This problem is amplified if the number of states increases. Hence, resampling steps are necessary and drastically improve efficiency and performance of particle MCMC samplers.
- However, posterior samples from particle Metropolis/Gibbs sampler still experience a high autocorrelation, and other, more advanced MCMC versions might yield better results.

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.

Chib, S. (1995). Marginal likelihood from the gibbs output. *Journal of the American Statistical Association*, 90(432):1313–1321.

Kantas, N., Doucet, A., Singh, S., Maciejowski, J., and Chopin, N. (2014). On particle methods for parameter estimation in general state-space models. *Statistical Science - Accepted for publication*.

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.

I would like to sincerely thank my supervisor Kostas Kalogeropoulos and co-supervisor Pauline Barrieu for their expertise, ideas, feedback and time.