Unbiased inference for discretely observed hidden Markov model diffusions

Jordan Franks

Department of Mathematics and Statistics, University of Jyväskylä, Finland

Bayesian computation for high-dimensional models
Closing workshop
IMS Singapore
19 September 2018
Joint work with Ajay Jasra, Kody Law and Matti Vihola

Hidden Markov model diffusion

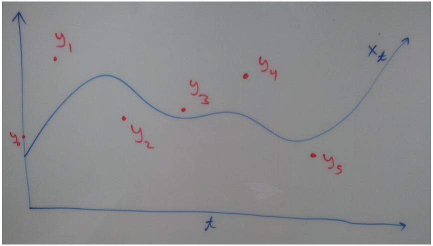
- Hidden process:

$$\mathrm{d}X'_t = a'^{(\theta)}_t(X'_t)\mathrm{d}t + b'^{(\theta)}_t(X'_t)\mathrm{d}W_t.$$

- Observations:

$$Y'_t|(\theta, X'_t) \sim g^{(\theta)}_t(Y'_t|X'_t).$$

We set $(X_p, Y_p) := (X'_{t_p}, Y'_{t_p})$, consider observations at times $\{t_p\}_{p=0}^n$.



Goal

Bayesian inference over joint posterior

$$\pi^{(\infty)}(\theta, x_{0:n}) \propto \mathrm{pr}(\theta) p_u^{(\theta,\infty)}(x_{0:n}) = \mathrm{pr}(\theta) L^{(\infty)}(\theta) p^{(\theta,\infty)}(x_{0:n}),$$

with $\mathrm{pr}(\theta)$ a prior density on model parameters, and

$$p^{(\theta,\infty)}(x_{0:n}) = \mathbb{P}_\theta[X_{0:n} = x_{0:n} | Y_{0:n} = y_{0:n}] = \frac{1}{L^{(\infty)}(\theta)} \left(\prod_{p=0}^{n} \underbrace{G_p^{(\theta)}(x_p)}_{:=g_{t_p}^{(\theta)}(y_p|x_p)} \right) \underbrace{\eta_0^{(\theta)}(x_0)}_{\eta_0^{(\theta)} \in \mathrm{Prob}(\mathbf{X})} \prod_{p=1}^{n} \underbrace{M_p^{(\theta,\infty)}(x_{p-1}, x_p)}_{\text{Markov kernel on } \mathbf{X}}.$$

- Challenges for unbiased, efficient and user-friendly inference
  - $M_p^{(\theta,\infty)}$ can not be evaluated or simulated.
  - n may be large (and $\mathbf{X}$ high-dimensional).
- Our approach:

Euler approximations + debiased MLMC + IS type correction.

Euler approximations

With $h_\ell = 2^{-\ell}$, $\ell \in \mathbb{N} \cup \{0\}$, new (appoximate) hidden Markov chain

$$X'^{(\ell)}_{t+h_\ell} = X'^{(\ell)}_t + a'^{(\theta)}_t(X'_t)h_\ell + b'^{(\theta)}_t(X'_t)(\underbrace{W_{t+h_\ell} - W_t}_{\sim\mathcal{N}(0,h_\ell)})$$

$\rightsquigarrow$ (approximate) Markov transition

$$M_p^{(\theta,\ell)}(x_{p-1}, x_p).$$

$\rightsquigarrow$ (approximate) latent probability

$$p^{(\theta,\ell)}(x_{0:n}).$$

$\rightsquigarrow$ (approximate) joint inference over $\pi^{(\theta,\ell)}$ possible (many papers) using

particle marginal Metropolis-Hastings (PMMH) (Andrieu, Doucet and Holenstein, "Particle MCMC" *JRSS-B* 2010).
Also: (Jasra, Kamatani, Law and Zhou, "Bayesian static parameter estimation..." *SIAM*, 2018).

Multilevel Monte Carlo via a coupling

MLMC approach used here: let $\phi : \mathbf{X}^{n+1} \to \mathbb{R}$ be a function.

- With $p_u^{(\theta,\ell)}(\phi) := L^{(\ell)}(\theta) p^{(\theta,\ell)}(\phi)$ approximate multilevel increment

$$p_u^{(\theta,\ell)}(\phi) - p_u^{(\theta,\ell-1)}(\phi)$$

  of *unnormalised* latent probability expectations $\rightsquigarrow$ "ΔPF."
  ΔPF in turn based on "approximate coupling" approach, at the Feynman-Kac model level, of (Jasra, Kamatani, Law, and Zhou, "Bayesian static parameter estimation..." *SIAM*, 2018):

  - With $\check{X}_p := (X_p^{(\ell)}, X_p^{(\ell-1)})$, set e.g.

$$\check{M}_p^{(\theta,\ell)}(\check{x}_{p-1}, \check{x}_p) := M_p^{(\theta,\ell)}(x_{p-1}^{(\ell)}, x_p^{(\ell)})\; M_p^{(\theta,\ell-1)}(x_{p-1}^{(\ell-1)}, x_p^{(\ell-1)})$$

    and set e.g.

$$\check{G}_p^{(\theta)}(\check{x}_p) = \frac{1}{2}\Big(G_p^{(\theta)}(x_p^{(\ell)}) + G_p^{(\theta)}(x_p^{(\ell-1)})\Big).$$

  - Then run standard particle filter (PF) for $(\check{M}_p, \check{G}_p)$.

Recall: particle filter for (M_p, G_p) with N particles

With $i \in \{1, \ldots N\}$, do:

- For $p = 0$, draw $X_0^{(i)} \sim \eta_0(\cdot)$.
  Set $V_0^{(i)} := \frac{1}{N} G_0(X_0^{(i)})$ and $V_0^* := \sum_{i=1}^N V_0^{(i)}$.
- For $1 \leq p \leq n$, draw $X_p^{(i)} \sim \sum_{j=1}^N \frac{G_{p-1}(X_{p-1}^{(j)}) M_p(X_{p-1}^{(j)}, \cdot)}{\sum_{k=1}^N G_{p-1}(X_{p-1}^{(k)})}$.
  Set $V_p^{(i)} := \left(V_{p-1}^*\right) \frac{1}{N} G_p(X_p^{(i)})$ and $V_p^* := \sum_{i=1}^N V_p^{(i)}$.

Output: $(X_{0:n}^{(i)}, V_n^{(i)})_{i=1}^N$.

The PF output satisfies for $\phi : \mathbf{X}^{n+1} \to \mathbb{R}$ (θ is fixed),

$$\mathbb{E}\Big[\sum_{i=1}^N V_n^{(i)} \phi(X_{0:n}^{(i)})\Big] = p_u^{(\theta)}(\phi) = L(\theta) p^{(\theta)}(\phi)$$

and

$$\mathbb{E}[V_n^*] = L(\theta).$$

The ΔPF for unnormalised level difference estimation

ΔPF algorithm is as follows:

1. Run the PF for $(\check{M}_p^{(\theta,\ell)}, \check{G}_p^{(\theta,\ell)})$, outputting $(\check{X}_{0:n}^{(i)}, V_n^{(i)})_{i=1}^N$.
2. Set

$$\Delta_\ell(\phi) := \sum_{i=1}^{N} V_n^{(i)} \Big(\overline{w}_\ell(\check{X}_{0:n}^{(i)}) \phi(\check{X}_{0:n}^{(\ell,i)}) - \underline{w}_\ell(\check{X}_{0:n}^{(i)}) \phi(\check{X}_{0:n}^{(\ell-1,i)}) \Big)$$

where

$$\overline{w}_\ell(\check{X}_p) := \frac{\prod_{p=0}^{n} G_p^{(\theta)}(\check{X}_p^{(\ell)})}{\prod_{p=0}^{n} \check{G}_p^{(\theta)}(\check{X}_p)} \qquad \text{and} \qquad \underline{w}_\ell(\check{X}_p) := \frac{\prod_{p=0}^{n} G_p^{(\theta)}(\check{X}_p^{(\ell-1)})}{\prod_{p=0}^{n} \check{G}_p^{(\theta)}(\check{X}_p)}.$$

Then

$$\begin{aligned}
&\mathbb{E}[\Delta_\ell(\phi)] \\
&= \int \phi(\check{x}_{0:n}^{(\ell)}) \Big(\prod_{p=0}^{n} G_p(\check{x}_p^{(\ell)}) \Big) \eta_0^{(\theta)} \prod_{p=1}^{n} \check{M}_p(\check{X}_{p-1}, \mathrm{d}\check{X}_p) - \int \text{'}(\ell-1)\text{-version'} \\
&= p_u^{(\theta,\ell)}(\phi) - p_u^{(\theta,\ell-1)}(\phi).
\end{aligned}$$

Debiased MLMC

McLeish (2011), Rhee and Glynn (2015), also cf. Vihola (2018).
We use single-term debiased MLMC estimator, i.e.
We randomise the running level of the ΔPF, i.e.

$$\Delta_\ell(\phi) \leftrightarrow p_L^{-1}\Delta_L(\phi)$$

That is,

- $(p_\ell)_{\ell\in\mathbb{N}}$ probability on $\mathbb{N}$ with $p_\ell > 0$ for all ℓ.
- Let $L \sim (p_\ell)$. Then if $\mathrm{var}(\Delta_L) < \infty$,

$$\mathbb{E}\Big[\frac{1}{p_L}\Delta_L(\phi)\Big] = p_u^{(\theta,\infty)}(\phi) - p_u^{(\theta,0)}(\phi).$$

$\implies$ We can target $p_u^{(\theta,\infty)}(\phi)$ by using

$$\bar{\Delta}_L^{(\theta)}(\phi) := p_L^{-1}\Delta_L(\phi) + \hat{p}_u^{(\theta,0)}(\phi),$$

where the term $\hat{p}_u^{(\theta,0)}(\phi) := \sum_{i=1}^N V_n^{(i)}\phi(X_{0:n}^{(\theta,0,i)})$ is simply formed from output of basic PF run for zeroth level model $(M_p^{(\theta,0)}, G_p^{(\theta)})$.

Unbiased joint inference

- The above implies unbiased (latent) inference over $p^{(\theta,\infty)}(x_{0:n})$.
- Recall: we want, more generally, unbiased joint inference over

$$\pi^{(\infty)}(\theta, x_{0:n}) \propto \mathrm{pr}(\theta) p_u^{(\theta)}(x_{0:n}) = \mathrm{pr}(\theta) L^{(\infty)}(\theta) p^{(\theta,\infty)}(x_{0:n}).$$

$\rightarrow$ we use an IS type correction of approximate MCMC, as suggested in (Vihola, Helske and F, "Importance sampling type estimators..." *arXiv* 2016): with $f : \mathbf{\Theta} \times \mathbf{X}^{n+1} \to \mathbb{R}$, $\epsilon \geq 0$ arbitrarily chosen,

(Phase 1) PMMH type chain $(\Theta_k, X_{0:n,k}^{(\Theta_k,0,i)})_{k\geq 1}^{i\in 1:N}$, with likelihood type estimates $(V_{n,k}^* + \epsilon)_{k\geq 1}$, targeting $\mathrm{pr}(\theta)(\hat{p}_u^{(\theta,0)}(x_{0:n}) + \epsilon)$.

(Phase 2) IS type correction based on debiased MLMC: $L_k \sim (p_\ell)$ and $\xi_k(\phi) := \frac{\bar{\Delta}_{L_k,k}^{(\Theta_k)}(\phi)}{(V_{n,k}^* + \epsilon)}$ for all $k \in 1{:}m$. The resulting estimator is

$$E_m^{IS}(f) = \frac{\sum_{k=1}^m \xi_k(f^{(\Theta_k)})}{\sum_{k=1}^m \xi_k(1)} \xrightarrow{m\to\infty} \pi^{(\infty)}(f),$$

with $f^{(\theta)}(x_{0:n}) := f(\theta, x_{0:n})$; consistency guaranteed if $\epsilon > 0$, PMMH ψ-irreducible, 1 and $f \in L^2(\pi^{(\infty)})$, and $\mathrm{pr}(p_L^{-1}\Delta_L(\phi)) < \infty$ for $\phi \in \{1, f^{(\cdot)}\}$.

Variance bound for the ΔPF

...suffices for the last condition to show uniform boundedness in θ of $s_\phi(\theta) := \sum_{\ell \geq 1} \frac{\mathbb{E}[\Delta_\ell^2(\phi)]}{p_\ell}$. $\implies$ we need a decay of ΔPF variance.

Theorem

Under standard conditions on the

- *diffusion (i.e. ellipticity, Lipschitz, bounded moments), and*
- *HMM/Feynman-Kac model (i.e. upper and lower bounded potentials, Lipschitz potentials and dynamics),*

for a function $\phi : \mathbf{X}^{n+1} \to \mathbb{R}$, bounded and Lipschitz, there is a constant $C = C(\phi, n)$ not depending on N, ℓ or θ, such that

$$\mathrm{var}\big(\Delta_\ell(\phi)\big) \leq C \frac{2^{-\beta\ell}}{N}$$

with $\beta = 1$ (Euler) or $\beta = 2$ (Milstein). Moreover,

$$\mathbb{E}\Big[\big(\Delta_\ell(\phi)\big)^2\Big] \leq C\Big(\frac{2^{-\beta\ell}}{N} + 2^{-2\ell}\Big).$$

Debiased MLMC cost and optimal allocations

How to choose (p_ℓ)? Answer more involved if consider computational costs too. We follow Glynn and Whitt (1992), Rhee and Glynn (2015).
$\rightsquigarrow$ Cost to run IS-ΔPF algorithm with m chain iterations:

$$\mathcal{C}(m) = \sum_{k=1}^{m} \tau_{\Theta_k, L_k}$$

Assume $\{\tau_{\Theta_k, L_k}\}_{k \geq 1}$ $\mathbb{R}_+$-random variables, conditionally independent given $\{\Theta_K\}$, satisfying for some $\gamma > 0$ (with constant $C > 0$ like before)

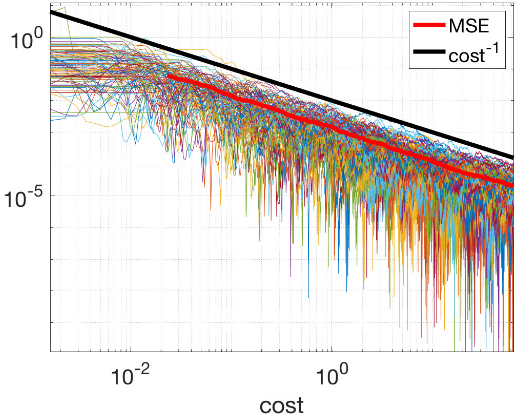
$$\mathbb{E}[\tau_{\Theta_k, L_k} | \Theta_k = \theta, L_k = \ell] \leq C 2^{\gamma \ell}.$$

Recommended allocations for particles N_ℓ and probability (p_ℓ):

- $(1 < \beta \leq 2$, e.g. Milstein $\beta = 2)$ Use $(N_\ell, p_\ell) \propto (1, 2^{-r\ell})$ for some $r \in (\gamma, \beta)$.
  - $\rightarrow$ canonical $\sqrt{m}$-convergence rate.
  - $\rightarrow$ e.g. with $(\beta, \gamma) = (2, 1)$ (typical 1D Milstein), use $p_\ell \propto 2^{-1.5\ell}$.
- $(\beta = 1$, Euler) for $\rho \in [0, 1]$, use $N_\ell \propto 2^{\rho \ell}$ and $p_\ell \propto 2^{-(1+\rho)\ell}$.
  - $\rightarrow$ Complexity $O(\epsilon^{-2} (\log \epsilon)^{2+\delta})$, $\delta > 0$ arbitrary.
  - $\rightarrow$ e.g. $(N_\ell, p_\ell) \propto (1, 2^{-\ell})$ or $\propto (2^\ell, 2^{-2\ell})$.

Ornstein-Uhlenbeck process (with Milstein)

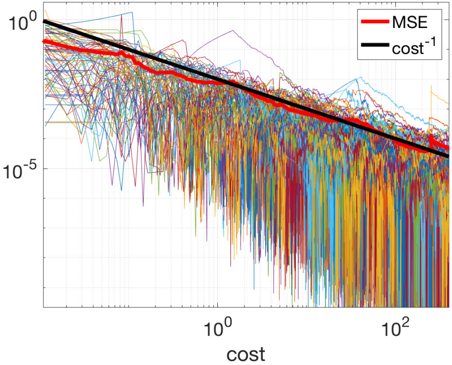
$$\mathrm{d}X_t = \theta_1 X_t \mathrm{d}t + \theta_2 \mathrm{d}W_t$$
$$Y_p = \mathcal{N}(X_p, 1), \qquad p = 1, \ldots, n = 5.$$



Drift-alternating geometric Brownian motion (with Euler)

$$\mathrm{d}X_{t+p} = \begin{cases} \theta^2 X_{t+p}\mathrm{d}t + \theta X_{t+p}\mathrm{d}W_{t+p}, & \text{if } p \text{ is even,} \\ \theta X_{t+p}\mathrm{d}W_{t+p}, & \text{else,} \end{cases} \quad \text{for } t \in (0,1),$$

$$Y_p \sim \mathcal{N}(\log(X_p), 1) \qquad p = 1, \ldots, n = 5.$$



Conclusion and further work

The IS algorithm for HMM diffusions:

- more broadly applicable, relying on Euler type approximations
- can provide unbiased/(asymptotically) exact inference
- efficient, achieving debiased MLMC complexity but in joint inference setting
- parallelisable (other IS efficiency enhancements possible)
- straightforward and easy to implement, and
- does not require tailoured solutions

A few ideas for possible investigations:

- extension to Lévy jump-diffusions, as in (Jasra, Law, Osei 2018)
- use of other approximate MCMC besides PMMH
- applying IS-debiasing technique outside diffusion context
- further efficiency considerations in practical situations and with limited computational resources, considering tail behaviour
- further stability and optimisation results for ΔPF and IS algorithm
- comparison of benefit of IS-debiasing with existing algorithms

Short literature list

C. Andrieu, A. Doucet, and R. Holenstein.
Particle Markov chain Monte Carlo methods.
J. R. Stat. Soc. Ser. B Stat. Methodol., 72(3):269–342, 2010.

A. Beskos, O. Papaspiliopoulos, G. O. Roberts, and P. Fearnhead.
Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes.
J. R. Stat. Soc. Ser. B Stat. Methodol., 68(3):333–382, 2006.

J. Franks, A. Jasra, K. Law and M. Vihola.
Unbiased inference for discretely observed hidden Markov model diffusions.
Preprint arXiv:1807.10259, 2018.

A. Jasra, K. Kamatani, K. Law, and Y. Zhou.
Bayesian static parameter estimation for partially observed diffusions via multilevel Monte Carlo.
SIAM J. Sci. Comp., 40:A887–A902, 2018.

A. Jasra, K. Law and P. Osei.
Multilevel particle filters for Lévy-driven stochastic differential equations.
Preprint arXiv:1804.04444, 2018.

C.-H. Rhee and P. W. Glynn.
Unbiased estimation with square root convergence for SDE models.
Oper. Res., 63(5):1026–1043, 2015.

G. Sermaidis, O. Papaspiliopoulos, G. Roberts, A. Beskos, and P. Fearnhead.
Markov chain Monte Carlo for exact inference for diffusions.
Scand. J. Statist, 40(2):294–321, 2013.

M. Vihola, J. Helske, and J. Franks.
Importance sampling type estimators based on approximate marginal MCMC.
Preprint arXiv:1609.02541v4, 2016.