Limit theorems for sequential MCMC methods (arXiv:1807.01057)

Axel Finke (Singapore) **Arnaud Doucet** (Oxford) **Adam M. Johansen** (Warwick)

Path-space Feynman–Kac models

Setup & notation:

- *path-space* formulation: $\mathbf{x}_n := x_{1:n} = (\mathbf{x}_{n-1}, x_n) \in \mathbf{E}_n := \mathbf{E}_{n-1} \times E$,
- Markov kernels: $M_n(\mathbf{x}_{n-1}, \mathrm{d}x_n)$,
- bounded potential functions: $G_n(\mathbf{x}_n) \in (0,1]$.

Goal: approximate distributions $(\eta_n)_{n\geq 1}$ on $(\mathbf{E}_n)_{n\geq 1}$:

$$\eta_n(\mathrm{d}\mathbf{x}_n) \propto \gamma_n(\mathrm{d}\mathbf{x}_n) := \eta_1 Q_{1,n}(\mathrm{d}\mathbf{x}_n),$$

$$Q_{p,n}(\mathrm{d}\mathbf{x}_n)(\mathbf{x}_p) := \prod_{q=p}^{n} G_{q-1}(\mathbf{x}_{q-1}) M_q(\mathbf{x}_{q-1}, \mathrm{d}x_q),$$

- unknown normalising constant: $\mathcal{Z}_n := \gamma_n(\mathbf{1})$,
- recursive definition: $\eta_n = \Phi_n^{\eta_{n-1}}$, where

$$\Phi_n^{\mu}(\mathrm{d}\mathbf{x}_n) := \frac{G_{n-1}(\mathbf{x}_{n-1})}{\mu(G_{n-1})}\mu(\mathrm{d}\mathbf{x}_{n-1}) M_n(\mathbf{x}_{n-1}, \mathrm{d}x_n).$$

Standard particle filter (PF) & MCMC-PF

Algorithm (PF). At time $n > 1$, given $\eta_{n-1}^N := \frac{1}{N}\sum_{i=1}^N \delta_{\boldsymbol{\xi}_{n-1}^i}$,

1. independently sample $\boldsymbol{\xi}_n^1, \dots, \boldsymbol{\xi}_n^N$ from
$$\Phi_n^{\eta_{n-1}^N}(\mathrm{d}\mathbf{x}_n) = \sum_{i=1}^N \frac{G_{n-1}(\boldsymbol{\xi}_{n-1}^i)}{\sum_{j=1}^N G_{n-1}(\boldsymbol{\xi}_{n-1}^j)} \delta_{\boldsymbol{\xi}_{n-1}^i}(\mathrm{d}\mathbf{x}_{n-1}) M_n(\mathbf{x}_{n-1}, \mathrm{d}x_n),$$
2. approximate η_n by $\eta_n^N := \frac{1}{N}\sum_{i=1}^N \delta_{\boldsymbol{\xi}_n^i}$.

Algorithm (MCMC-PF). At time $n > 1$, given $\eta_{n-1}^N := \frac{1}{N}\sum_{i=1}^N \delta_{\boldsymbol{\xi}_{n-1}^i}$,

1. initialise $\boldsymbol{\xi}_n^1 \sim \kappa_n^{\eta_{n-1}^N} \approx \Phi_n^{\eta_{n-1}^N}$,
2. for $2 \leq i \leq N$, sample
$$\boldsymbol{\xi}_n^i \sim \underbrace{K_n^{\eta_{n-1}^N}}_{\Phi_n^{\eta_{n-1}^N}\text{-invariant MCMC kernel}}(\boldsymbol{\xi}_n^{i-1}, \cdot\,),$$
3. approximate η_n by $\eta_n^N := \frac{1}{N}\sum_{i=1}^N \delta_{\boldsymbol{\xi}_n^i}$.

- Versions of MCMC-PFs a.k.a. *sequential MCMC* methods proposed in Berzuini et al. (1997); Septier and Peters (2016).
- Reduce to standard PFs if $K_n^{\mu}(\mathbf{x}_n, \cdot\,) \equiv \Phi_n^{\mu} = \kappa_n^{\mu}$.
- Usual estimates of $\gamma_n(\varphi_n)$ and $\mathcal{Z}_n$:

$$\gamma_n^N(\varphi_n) := \eta_n^N(\varphi_n) \prod_{p=1}^{n-1} \eta_p^N(G_p),$$

$$\mathcal{Z}_n^N := \gamma_n^N(\mathbf{1}) = \prod_{p=1}^{n-1} \frac{1}{N}\sum_{i=1}^N G_p(\boldsymbol{\xi}_p^i).$$

Application to state-space models

Example (bootstrap PF (BPF)-type flow).

$$G_n(\mathbf{x}_n) := g(y_n|x_n),$$
$$M_n(\mathbf{x}_{n-1}, \mathrm{d}x_n) := f(\mathrm{d}x_n|x_{n-1}).$$

In this case, $\eta_n(\mathrm{d}\mathbf{x}_n) = p(\mathrm{d}x_{1:n}|y_{1:n-1})$, $\mathcal{Z}_{n+1} = p(y_{1:n})$ and

$$\Phi_n^{\eta_{n-1}^N}(\mathrm{d}\mathbf{x}_n) = \sum_{i=1}^N \frac{g(y_{n-1}|\xi_{n-1}^i)}{\sum_{j=1}^N g(y_{n-1}|\xi_{n-1}^j)} \times \delta_{\boldsymbol{\xi}_{n-1}^i}(\mathrm{d}\mathbf{x}_{n-1}) f(\mathrm{d}x_n|\xi_{n-1}^i).$$

$\Rightarrow$ can typically implement both BPF and MCMC-BPF.

Example (fully-adapted auxiliary PF (FA-APF)-type flow).

$$G_n(\mathbf{x}_n) := p(y_{n+1}|x_n), \leftarrow \text{typically intractable!}$$
$$M_n(\mathbf{x}_{n-1}, \mathrm{d}x_n) := p(\mathrm{d}x_n|y_n, x_{n-1}) := \frac{g(y_n|x_n) f(\mathrm{d}x_n|x_{n-1})}{p(y_n|x_{n-1})}.$$

In this case, $\eta_n(\mathrm{d}\mathbf{x}_n) = p(\mathrm{d}x_{1:n}|y_{1:n})$, $\mathcal{Z}_n = p(y_{1:n})$ and

$$\Phi_n^{\eta_{n-1}^N}(\mathrm{d}\mathbf{x}_n) = \sum_{i=1}^N \frac{p(y_n|\xi_{n-1}^i)}{\sum_{j=1}^N p(y_n|\xi_{n-1}^j)} \times \delta_{\boldsymbol{\xi}_{n-1}^i}(\mathbf{x}_{n-1}) p(\mathrm{d}x_n|y_n, \xi_{n-1}^i)$$
$$\propto \sum_{i=1}^N g(y_n|x_n) \times \delta_{\boldsymbol{\xi}_{n-1}^i}(\mathrm{d}\mathbf{x}_{n-1}) f(\mathrm{d}x_n|\xi_{n-1}^i).$$

$\Rightarrow$ can implement MCMC-FA-APF even if FA-APF cannot be used.

- More general auxiliary PF flows could be considered.

Analysis of MCMC-PFs

- Assumptions (informally stated; similar to Bercu et al. (2012)):

A1 Uniform ergodicity: K_n^{μ} is uniformly geometrically ergodic, uniformly in $\mu \in \mathcal{P}(\mathbf{E}_{n-1})$.

A2 Lipschitz property: we can suitably control $\|K_n^{\mu} - K_n^{\nu}\|$ by controlling $\|\mu - \nu\|$.

Proposition (unbiasedness). For any $n \geq 1$, $N \geq 1$ and $\varphi_n \in \mathcal{B}(\mathbf{E}_n)$, if the chains are initialised from stationarity, i.e. if $\kappa_p^{\mu} = \Phi_p^{\mu}$, for $1 \leq p \leq n$,

1. $\mathbb{E}[\gamma_n^N(\varphi_n)] = \gamma_n(\varphi_n)$,
2. $\mathbb{E}[\mathcal{Z}_n^N] = \mathcal{Z}_n$.

Proposition (strong law of large numbers). Under **A1**, for any $n \geq 1$ and $\varphi_n \in \mathcal{B}(\mathbf{E}_n)$, as $N \to \infty$,

1. $\gamma_n^N(\varphi_n) \to_{\text{a.s.}} \gamma_n(\varphi_n)$,
2. $\eta_n^N(\varphi_n) \to_{\text{a.s.}} \eta_n(\varphi_n)$.

Analysis of MCMC-PFs, continued

For any ν-invariant Markov kernel K, define the integrated autocorrelation time

$$\mathrm{iact}_K[\varphi] := 1 + 2\sum_{l=1}^{\infty} \frac{\mathrm{cov}_{\nu}[\varphi, K^l(\varphi)]}{\mathrm{var}_{\nu}[\varphi]}.$$

Proposition (central limit theorem). Under **A1** and **A2**, for any $n \geq 1$ and any $\varphi_n \in \mathcal{B}(\mathbf{E}_n)$,

1. $\sqrt{N}[\gamma_n^N/\gamma_n(\mathbf{1}) - \eta_n](\varphi_n) \to_{\mathrm{d}} \mathrm{N}(0, \sigma_n^2[\varphi_n])$,
2. $\sqrt{N}[\eta_n^N - \eta_n](\varphi_n) \to_{\mathrm{d}} \mathrm{N}(0, \sigma_n^2[\varphi_n - \eta_n(\varphi_n)])$,

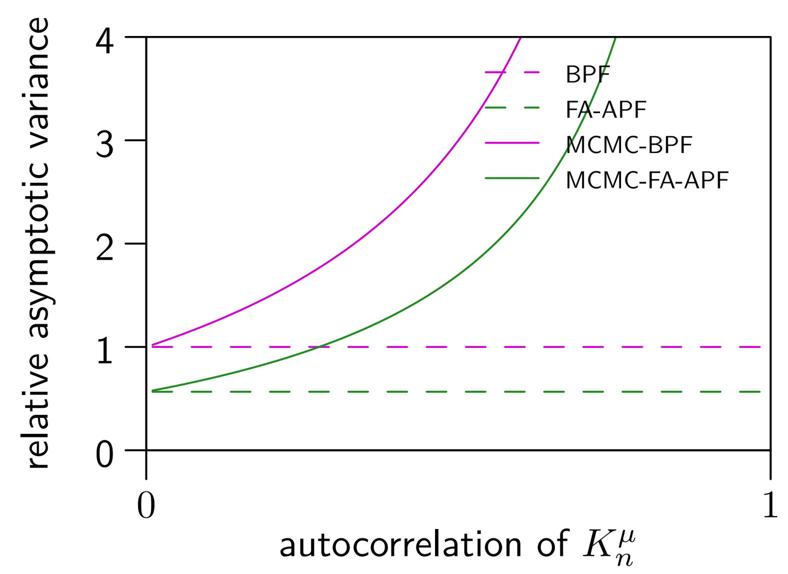
as $N \to \infty$, with asymptotic variance

$$\sigma_n^2[\varphi] := \sum_{p=1}^{n} \mathrm{var}_{\eta_p}[\bar{Q}_{p,n}(\varphi)] \times \overbrace{\mathrm{iact}_{K_p^{\eta_{p-1}}}[\bar{Q}_{p,n}(\varphi)]}^{\substack{\text{(typically)} > 1 \text{ for MCMC-PFs} \\ = 1 \text{ for standard PFs}}}.$$

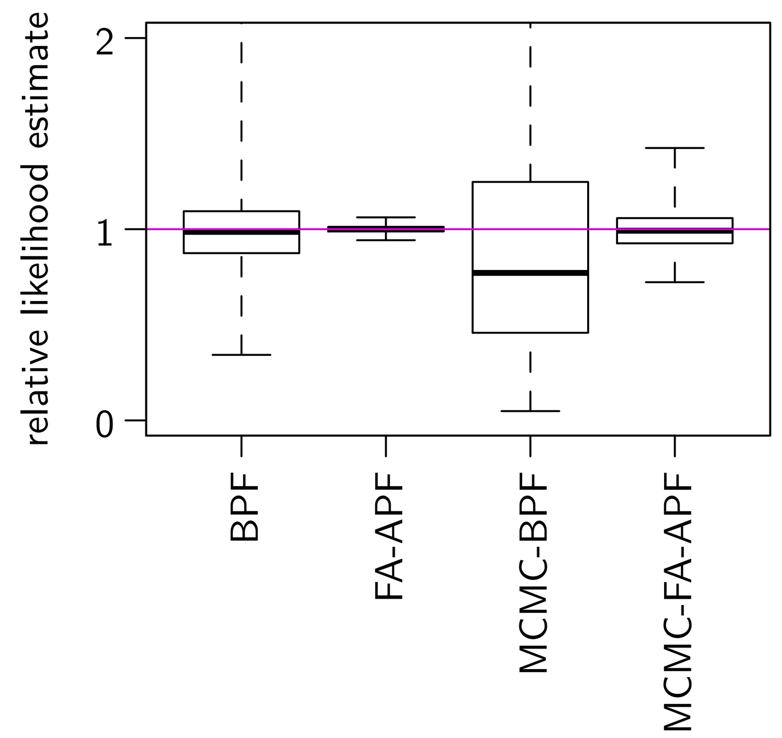
Here, $\bar{Q}_{p,n} := \frac{\gamma_n(\mathbf{1})}{\gamma_p(\mathbf{1})} Q_{p,n}$ satisfies $\eta_p \bar{Q}_{p,n} = \eta_n$.

- Under further (e.g. strong mixing) assumptions, the convergence rate/asymptotic variance is bounded uniformly in n.

Numerical illustration



Binary state-space model. Asymptotic variances (relative to the asymptotic variance of the BPF).



5-dimensional linear-Gaussian state-space model. Estimates of the marginal likelihood (relative to the true marginal likelihood) using $N = 10{,}000$ particles.

Literature

Bercu, B., Del Moral, P., and Doucet, A. (2012). Fluctuations of interacting Markov chain Monte Carlo methods. *Stochastic Processes and their Applications*, 122(4):1304–1331.

Berzuini, C., Best, N. G., Gilks, W. R., and Larizza, C. (1997). Dynamic conditional independence models and Markov chain Monte Carlo methods. *Journal of the American Statistical Association*, 92(440):1403–1412.

Septier, F. and Peters, G. W. (2016). Langevin and Hamiltonian based sequential MCMC for efficient Bayesian filtering in high-dimensional spaces. *IEEE Journal of Selected Topics in Signal Processing*, 10(2):312–327.