# Limit theorems for sequential MCMC methods

Axel Finke[1] Arnaud Doucet[2] Adam M. Johansen[3]

[1]DSAP, NUS

[2]Dept of Statistics, University of Oxford

[3]Dept of Statistics, University of Warwick; The Alan Turing Institute

4th September 2018

Standard PFs and MCMC-PFs

Convergence analysis of MCMC-PFs

Application to state-space models

Numerical illustrations

# Summary

- **Sequential MCMC** methods a.k.a. **MCMC-PFs:**
  - proposed in Berzuini et al. (1997), extended in Septier et al. (2009); Septier and Peters (2016); Finke et al. (2016).
- **Difference with (standard) *particle filters (PFs)*:**
  - PFs sample/resample particles conditionally independently,
  - MCMC-PFs sample/resample particles jointly according to a Markov chain.
- **This work:**
  - convergence analysis of MCMC-PFs;
  - guidance on when to use standard PFs/MCMC-PFs

# Path-space Feynman–Kac model

**Setup & notation:**

- *path-space* formulation:
  $\mathbf{x}_n := x_{1:n} = (\mathbf{x}_{n-1}, x_n) \in \mathbf{E}_n := \mathbf{E}_{n-1} \times E$,
- mutation 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 $(\mathbf{E}_n)_{n\geq 1}$:

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

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

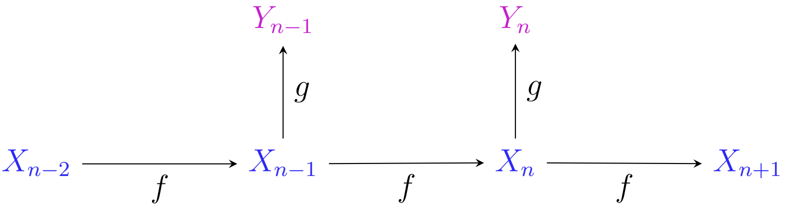
- unknown normalising constant: $\mathcal{Z}_n := \gamma_n(\mathbf{1})$,
- recursive definition: $\eta_n = \Phi_n^{\eta_{n-1}}$, where

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

# Example: Bootstrap PF flow I

**State-space model:**

- Bivariate Markov chain $(X_n, Y_n)_{n\in\mathbb{N}}$,
- with transition kernel $f(\mathrm{d}x_n|x_{n-1})g(y_n|x_n)\mathrm{d}y_n$,
- only $Y_n = y_n$ is observed; $X_n$ is latent.



# Example: Bootstrap PF flow II

Take

$$
\begin{aligned}
G_n(\mathbf{x}_n) &\coloneqq g(y_n|x_n), \\
M_n(\mathbf{x}_{n-1}, \mathrm{d}x_n) &\coloneqq f(\mathrm{d}x_n|x_{n-1}).
\end{aligned}
$$

Then

$$
\begin{aligned}
\eta_n(\mathrm{d}\mathbf{x}_n) &= p(\mathrm{d}x_{1:n}|y_{1:n-1}) \\
&= \Phi_n^{\eta_{n-1}}(\mathrm{d}\mathbf{x}_n) \\
&= \frac{\eta_{n-1}(\mathrm{d}\mathbf{x}_{n-1}) G_{n-1}(\mathbf{x}_{n-1})}{\eta_{n-1}(G_{n-1})} M_n(\mathbf{x}_{n-1}, \mathrm{d}x_n) \\
&= \frac{p(\mathrm{d}x_{1:n-1}|y_{1:n-2}) g(y_{n-1}|x_{n-1})}{\int p(\mathrm{d}x'_{1:n-1}|y_{1:n-2}) g(y_{n-1}|x'_{n-1})} f(\mathrm{d}x_n|x_{n-1}),
\end{aligned}
$$

as well as $\mathcal{Z}_n = p(y_{1:n-1})$.

- **Problem:** $\eta_n$ is intractable.
- **Idea:** recursively construct approximation $\eta_n^N$ of $\eta_n = \Phi_n^{\eta_{n-1}}$.
  1. given $\eta_{n-1}^N := \frac{1}{N}\sum_{i=1}^N \delta_{\boldsymbol{\xi}_{n-1}^i}$, obtain the mixture
  $$\Phi_n^{\eta_{n-1}^N} = \sum_{i=1}^N \frac{G_{n-1}(\boldsymbol{\xi}_{n-1}^i)}{\sum_{j=1}^N G_{n-1}(\boldsymbol{\xi}_{n-1}^j)} [\delta_{\boldsymbol{\xi}_{n-1}^i} \otimes M_n],$$
  2. sample $N$ *particles* $\boldsymbol{\xi}_n^1, \ldots, \boldsymbol{\xi}_n^N$ (approximately) from $\Phi_n^{\eta_{n-1}^N}$,
  3. approximate $\eta_n$ by $\eta_n^N := \frac{1}{N}\sum_{i=1}^N \delta_{\boldsymbol{\xi}_n^i}$.

**Algorithm (PF).** In Step 2, sample $\boldsymbol{\xi}_n^1, \ldots, \boldsymbol{\xi}_n^N \overset{\text{iid}}{\sim} \Phi_n^{\eta_{n-1}^N}$.

**Algorithm (MCMC-PF).** In Step 2,

- initialise $\boldsymbol{\xi}_n^1 \sim \kappa_n^{\eta_{n-1}^N} \approx \Phi_n^{\eta_{n-1}^N}$,
- sample $\boldsymbol{\xi}_n^i \sim \underbrace{K_n^{\eta_{n-1}^N}}_{\Phi_n^{\eta_{n-1}^N}\text{-invariant MCMC kernel}}(\boldsymbol{\xi}_n^{i-1}, \cdot\,)$, for $i = 2, \ldots, N$.

Standard PFs and MCMC-PFs

Convergence analysis of MCMC-PFs

Application to state-space models

Numerical illustrations

- Recall: $\eta_n(\mathrm{d}\mathbf{x}_n) = \gamma_n(\mathrm{d}\mathbf{x}_n)/\mathcal{Z}_n$.
- Usual estimates of $\gamma_n(\varphi_n)$ and $\mathcal{Z}_n$:

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

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

**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. $\mathbb{E}[\mathcal{Z}_n^N] = \mathcal{Z}_n$.

# Assumptions

**A1** For any $n \geq 1$, there exists $i_n \in \mathbb{N}$ such that

$$\sup_{\mu \in \mathcal{P}(\mathbf{E}_{n-1})} \beta((K_n^\mu)^{i_n}) < 1,$$

where $\beta(K) := \sup_{x,y} \|K(x,\cdot) - K(y,\cdot)\|$.

**A2** For any $n \geq 1$, there exists a constant $\bar{\Gamma}_n < \infty$ and a family of bounded integral operators $(\Gamma_n^\mu)_{\mu \in \mathcal{P}(\mathbf{E}_{n-1})}$ from $\mathcal{B}(\mathbf{E}_{n-1})$ to $\mathcal{B}(\mathbf{E}_n)$ s.t. for any $(\mu,\nu) \in \mathcal{P}(\mathbf{E}_{n-1})^2$ and any $f_n \in \mathcal{B}(\mathbf{E}_n)$,

$$\|[K_n^\mu - K_n^\nu](f_n)\| \leq \int_{\mathcal{B}(\mathbf{E}_{n-1})} |[\mu - \nu](g)| \Gamma_n^\mu(f_n, \mathrm{d}g)$$

and $\int_{\mathcal{B}(\mathbf{E}_{n-1})} \|g\| \Gamma_n^\mu(f_n, \mathrm{d}g) \leq \|f_n\| \bar{\Gamma}_n$.

- strong but similar to assumptions in Del Moral and Doucet (2010); Brockwell et al. (2010); Bercu et al. (2012),
- satisfied, e.g.if $K_n^\mu$ is an independent MH kernel & $E$ finite.

**Proposition ($\mathbb{L}_p$-error bound).** Under **A1**, for any $n, p \geq 1$, there exist $a_n, b_p < \infty$ such that for any $\varphi_n \in \mathcal{B}(\mathbf{E}_n)$ and any $N \geq 1$,

$$\mathbb{E}\left[\left|[\eta_n^N - \eta_n](\varphi_n)\right|^p\right]^{\frac{1}{p}} \leq \frac{a_n b_p}{\sqrt{N}} \|\varphi_n\|.$$

- Under strong mixing assumptions and if $\varphi_n(x_{1:n}) = \varphi(x_n)$, $\sup_{n \geq 1} a_n < \infty$.

**Corollary (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. $\eta_n^N(\varphi_n) \to_{\text{a.s.}} \eta_n(\varphi_n)$.

For any $\nu$-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–A2**, for any $n \geq 1$ and any $\varphi_n \in \mathcal{B}(\mathbf{E}_n)$, as $N \to \infty$,

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

with asymptotic variance

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

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 strong mixing assumptions and if $\varphi_n(x_{1:n}) = \varphi(x_n)$, $\sup_{n\geq 1} \sigma_n^2[\varphi_n - \eta_n(\varphi_n)] < \infty$.

Standard PFs and MCMC-PFs

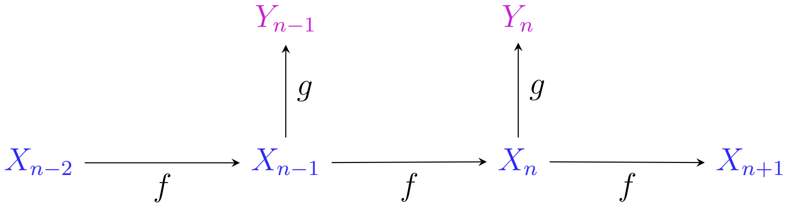
Convergence analysis of MCMC-PFs

Application to state-space models

Numerical illustrations

# State-space model

- Bivariate Markov chain $(X_n, Y_n)_{n \in \mathbb{N}}$,
- with transition kernel $f(\mathrm{d}x_n | x_{n-1}) g(y_n | x_n) \mathrm{d}y_n$,
- only $Y_n = y_n$ is observed; $X_n$ is latent.



# Bootstrap PF (BPF)-type flow

**Example (BPF flow).**

$$G_{n-1}(\mathbf{x}_{n-1}) \coloneqq g(y_{n-1}|x_{n-1}),$$
$$M_n(\mathbf{x}_{n-1}, \mathrm{d}x_n) \coloneqq f(\mathrm{d}x_n|x_{n-1}).$$

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

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

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

# Fully-adapted auxiliary PF (FA-APF)-type flow

**Example (FA-APF flow).**

$$G_{n-1}(\mathbf{x}_{n-1}) \coloneqq p(y_n|x_{n-1}) = \overbrace{\int g(y_n|x_n) f(\mathrm{d}x_n|x_{n-1})}^{\text{typically intractable!}}$$

$$M_n(\mathbf{x}_{n-1}, \mathrm{d}x_n) \coloneqq p(\mathrm{d}x_n|y_n, x_{n-1}) \coloneqq \frac{g(y_n|x_n) f(\mathrm{d}x_n|x_{n-1})}{p(y_n|x_{n-1})}.$$

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

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

$\Rightarrow$ can typically implement MCMC-FA-APF **but not** FA-APF.

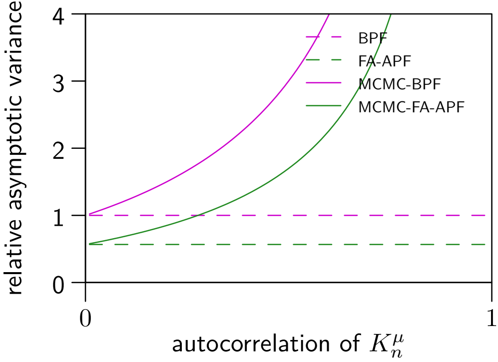
# Variance–variance trade-off

| | BPF flow | FA-APF flow |
|---|---|---|
| Standard PF | BPF | MCMC-FA-APF (usually intractable) |
| MCMC-PF | MCMC-BPF (not very useful) | MCMC-FA-APF |

- PFs preferable if they target the same distribution flow.
- MCMC-PFs preferable if
  - they can target a more efficient distribution flow,
  - the MCMC kernels do not mix too poorly.

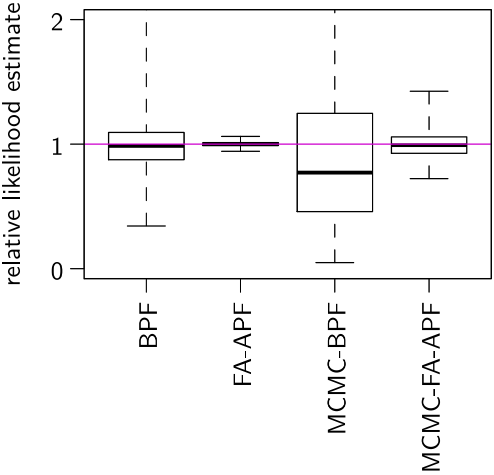
**Trade-off:** variance due to importance-sampling vs. variance due to additional particle (auto-)correlation.

# 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.

# Ongoing work

**With Alex Thiery:**

- additional dependence between particles may be more useful within *'conditional'* SMC algorithms,
- permits 'local' conditional SMC algorithms,
  - better scaling in high dimensions,
  - example: *embedded hidden Markov model* method (Shestopaloff and Neal, 2018) which is the conditional SMC version of MCMC-PFs.
- more on this in my talk next week.

# 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.

Brockwell, A., Del Moral, P., and Doucet, A. (2010). Sequentially interacting Markov chain Monte Carlo methods. *The Annals of Statistics*, 38(6):3387–3411.

Del Moral, P. and Doucet, A. (2010). Interacting Markov chain Monte Carlo methods for solving nonlinear measure-valued equations. *The Annals of Applied Probability*, 20(2):593–639.

Finke, A., Doucet, A., and Johansen, A. M. (2016). On embedded hidden Markov models and particle Markov chain Monte Carlo methods. *ArXiv e-prints*, 1610.08962.

Septier, F., Pang, S. K., Carmi, A., and Godsill, S. (2009). On MCMC-based particle methods for Bayesian filtering: Application to multitarget tracking. In *3rd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 360–363. IEEE.

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.

Shestopaloff, A. Y. and Neal, R. M. (2018). Sampling latent states for high-dimensional non-linear state space models with the embedded HMM method. *Bayesian Analysis*, 13(3):797–822.