

○
○○○
○○

○○
○○
○○○○

○○○○○○○
○○
○○
○

○

A Modern Perspective on Auxiliary Particle Filters

Adam M. Johansen and Nick Whiteley

`npw24@cam.ac.uk`

Isaac Newton Institute — 18th June 2008

Collaborators: Arnaud Doucet, Simon Godsill and Sumeet Singh.

○
○○○
○○

○○
○○
○○○○

○○○○○○○
○○
○○
○

○

Introduction



Outline

- ▶ Background: Particle Filters
 - ▶ Hidden Markov Models & Filtering
 - ▶ Particle Filters & Sequential Importance Resampling
 - ▶ Auxiliary Particle Filters
- ▶ Interpretation
 - ▶ Auxiliary Particle Filters are SIR Algorithms
 - ▶ Theoretical Considerations
 - ▶ Practical Implications
- ▶ Applications
 - ▶ Auxiliary SMC Samplers
 - ▶ The Probability Hypothesis Density
 - ▶ On Stratification
- ▶ Conclusions



Outline

- ▶ Background: Particle Filters
 - ▶ Hidden Markov Models & Filtering
 - ▶ Particle Filters & Sequential Importance Resampling
 - ▶ Auxiliary Particle Filters
- ▶ Interpretation
 - ▶ Auxiliary Particle Filters are SIR Algorithms
 - ▶ Theoretical Considerations
 - ▶ Practical Implications
- ▶ Applications
 - ▶ Auxiliary SMC Samplers
 - ▶ The Probability Hypothesis Density
 - ▶ On Stratification
- ▶ Conclusions



Outline

- ▶ Background: Particle Filters
 - ▶ Hidden Markov Models & Filtering
 - ▶ Particle Filters & Sequential Importance Resampling
 - ▶ Auxiliary Particle Filters
- ▶ Interpretation
 - ▶ Auxiliary Particle Filters are SIR Algorithms
 - ▶ Theoretical Considerations
 - ▶ Practical Implications
- ▶ Applications
 - ▶ Auxiliary SMC Samplers
 - ▶ The Probability Hypothesis Density
 - ▶ On Stratification
- ▶ Conclusions

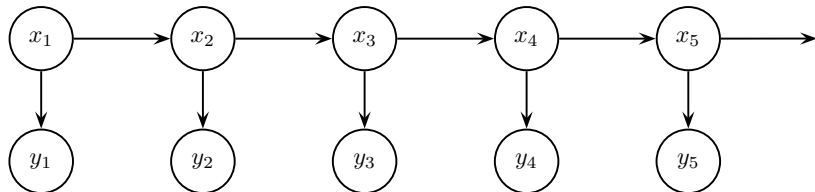


Outline

- ▶ Background: Particle Filters
 - ▶ Hidden Markov Models & Filtering
 - ▶ Particle Filters & Sequential Importance Resampling
 - ▶ Auxiliary Particle Filters
- ▶ Interpretation
 - ▶ Auxiliary Particle Filters are SIR Algorithms
 - ▶ Theoretical Considerations
 - ▶ Practical Implications
- ▶ Applications
 - ▶ Auxiliary SMC Samplers
 - ▶ The Probability Hypothesis Density
 - ▶ On Stratification
- ▶ Conclusions



Hidden Markov Models



- ▶ X_n is a \mathcal{X} -valued Markov Chain with transition density f :

$$X_n | \{X_{n-1} = x_{n-1}\} \sim f(\cdot | x_{n-1})$$

- ▶ Y_n is a \mathcal{Y} -valued stochastic process:

$$Y_n | \{X_n = x_n\} \sim g(\cdot | x_n)$$



Optimal Filtering

- ▶ The *filtering distribution* may be expressed recursively as:

$$p(x_n | y_{1:n-1}) = \int p(x_{n-1} | y_{1:n-1}) f(x_n | x_{n-1}) dx_{n-1} \quad \text{Prediction}$$

$$p(x_n | y_{1:n}) = \frac{p(x_n | y_{1:n-1}) g(y_n | x_n)}{\int p(x_n | y_{1:n-1}) g(y_n | x_n) dx_n} \quad \text{Update.}$$

- ▶ The *smoothing distribution* may be expressed recursively as:

$$p(x_{1:n} | y_{1:n-1}) = p(x_{1:n-1} | y_{1:n-1}) f_n(x_n | x_{n-1}) \quad \text{Prediction}$$

$$p(x_{1:n} | y_{1:n}) = \frac{p(x_{1:n} | y_{1:n-1}) g_n(y_n | x_n)}{\int p(x_{1:n} | y_{1:n-1}) g_n(y_n | x_n) dx_{1:n}} \quad \text{Update.}$$



Optimal Filtering

- ▶ The *filtering distribution* may be expressed recursively as:

$$p(x_n | y_{1:n-1}) = \int p(x_{n-1} | y_{1:n-1}) f(x_n | x_{n-1}) dx_{n-1} \quad \text{Prediction}$$

$$p(x_n | y_{1:n}) = \frac{p(x_n | y_{1:n-1}) g(y_n | x_n)}{\int p(x_n | y_{1:n-1}) g(y_n | x_n) dx_n} \quad \text{Update.}$$

- ▶ The *smoothing distribution* may be expressed recursively as:

$$p(x_{1:n} | y_{1:n-1}) = p(x_{1:n-1} | y_{1:n-1}) f_n(x_n | x_{n-1}) \quad \text{Prediction}$$

$$p(x_{1:n} | y_{1:n}) = \frac{p(x_{1:n} | y_{1:n-1}) g_n(y_n | x_n)}{\int p(x_{1:n} | y_{1:n-1}) g_n(y_n | x_n) dx_{1:n}} \quad \text{Update.}$$



Particle Filters: Sequential Importance Resampling

At time $n = 1$:

- ▶ Sample $X_{1,1}^i \sim q_1(\cdot)$.
- ▶ Weight

$$W_1^i \propto f(X_{1,1}^i)g(y_1|X_{1,1}^i)/q_1(X_{1,1}^i)$$

- ▶ Resample.

At times $n > 1$, iterate:

- ▶ Sample $X_{n,n}^i \sim q_n(\cdot|X_{n-1,n-1}^i)$. Set $X_{n,1:n-1}^i = X_{n-1}^i$.
- ▶ Weight

$$W_n^i \propto \frac{f(X_{n,n}^i|X_{n,n-1}^i)g(X_{n,n}^i|y_n)}{q_n(X_{n,n}^i|X_{n,1:n-1}^i)}$$

- ▶ Resample.



Auxiliary [v] Particle Filters (Pitt & Shephard '99)

If we have access to the next observation before resampling, we could use this structure:

- ▶ Pre-weight every particle with $\lambda_n^{(i)} \propto \hat{p}(y_n | X_{n-1}^{(i)})$.
- ▶ Propose new states, from the mixture distribution

$$\sum_{i=1}^N \lambda_n^{(i)} q(\cdot | X_{n-1}^{(i)}) / \sum_{i=1}^N \lambda_n^{(i)}.$$

- ▶ Weight samples, correcting for the pre-weighting.

$$W_n^i \propto \frac{f(X_{n,n}^i | X_{n,n-1}^i) g(X_{n,n}^i | y_n)}{\lambda_n^i q_n(X_{n,n}^i | X_{n,1:n-1}^i)}$$

- ▶ Resample particle set.



Some Well Known Refinements

We can tidy things up a bit:

1. The auxiliary variable step is equivalent to multinomial resampling.
2. So, there's no need to resample before the pre-weighting.

Now we have:

- ▶ Pre-weight every particle with $\lambda_n^{(i)} \propto \hat{p}(y_n | X_{n-1}^{(i)})$.
- ▶ Resample
- ▶ Propose new states
- ▶ Weight samples, correcting for the pre-weighting.

$$W_n^i \propto \frac{f(X_{n,n}^i | X_{n,n-1}^i) g(X_{n,n}^i | y_n)}{\lambda_n^i q_n(X_{n,n}^i | X_{n,1:n-1})}$$

o
oo
oo

oo
oo
oooo

oooooo
oo
oo
o

o

Interpretation



General SIR Algorithms

The SIR algorithm can be used somewhat more generally.

Given $\{\pi_n\}$ defined on $E_n = \mathcal{X}^n$:

- ▶ Sample $X_{n,n}^i \sim q_n(\cdot | X_{n-1}^i)$. Set $X_{n,1:n-1}^i = X_{n-1}^i$.
- ▶ Weight

$$W_n^i \propto \frac{\pi_n(X_n^i)}{\pi_{n-1}(X_{n,1:n-1}^i) q_n(X_{n,n}^i | X_{n,1:n-1}^i)}$$

- ▶ Resample.



An Interpretation of the APF

If we move the first step at time $n + 1$ to the last at time n , we get:

- ▶ Resample
- ▶ Propose new states
- ▶ Weight samples, correcting earlier pre-weighting.
- ▶ Pre-weight every particle with $\lambda_{n+1}^{(i)} \propto \hat{p}(y_{n+1}|X_n^{(i)})$.

which is an SIR algorithm targetting the sequence of distributions

$$\eta_n(x_n) \propto p(x_{1:n}|y_{1:n})\hat{p}(y_{n+1}|x_n)$$

which allows estimation under the actually interesting distribution via importance sampling.



Theoretical Considerations

- ▶ Direct analysis of the APF is largely unnecessary.
- ▶ Results can be obtained by considering the associated SIR algorithm.
- ▶ SIR has a (discrete time) Feynman-Kac interpretation.



For example...

Proposition. Under standard regularity conditions

$$\sqrt{N} (\hat{\varphi}_{n,APF}^N - \bar{\varphi}_n) \rightarrow \mathcal{N}(0, \sigma_n^2(\varphi_n))$$

where,

$$\sigma_1^2(\varphi_1) = \int \frac{p(x_1|y_1)^2}{q_1(x_1)} (\varphi_1(x_1) - \bar{\varphi}_1)^2 dx_1$$

$$\begin{aligned} \sigma_n^2(\varphi_n) &= \int \frac{p(x_1|y_{1:n})^2}{q_1(x_1)} \left(\int \varphi_n(x_{1:n}) p(x_{2:n}|y_{2:n}, x_1) dx_{2:n} - \bar{\varphi}_n \right)^2 dx_1 \\ &+ \sum_{k=2}^{t-1} \int \frac{p(x_{1:k}|y_{1:n})^2}{\hat{\mathbf{p}}(\mathbf{x}_{1:k-1}|\mathbf{y}_{1:k}) q_k(x_k|x_{k-1})} \left(\int \varphi_n(x_{1:n}) p(x_{k+1:n}|y_{k+1:n}, x_k) dx_{k+1:n} - \bar{\varphi}_n \right)^2 dx_{1:k} \\ &+ \int \frac{p(x_{1:n}|y_{1:n})^2}{\hat{\mathbf{p}}(\mathbf{x}_{1:n-1}|\mathbf{y}_{1:n}) q_n(x_n|x_{n-1})} (\varphi_n(x_{1:n}) - \bar{\varphi}_n)^2 dx_{1:n}. \end{aligned}$$



Practical Implications

- ▶ It means we're doing importance sampling.
- ▶ Choosing $\widehat{p}(y_n|x_{n-1}) = p(y_n|x_n = \mathbb{E}[X_n|x_{n-1}])$ is dangerous.
- ▶ A safer choice would be ensure that

$$\sup_{x_{n-1}, x_n} \frac{g(y_n|x_n)f(x_n|x_{n-1})}{\widehat{p}(y_n|x_{n-1})q(x_n|x_{n-1})} < \infty$$

- ▶ Using APF doesn't *ensure* superior performance.



A Contrived Illustration

Consider the following binary state-space model with common state and observation spaces:

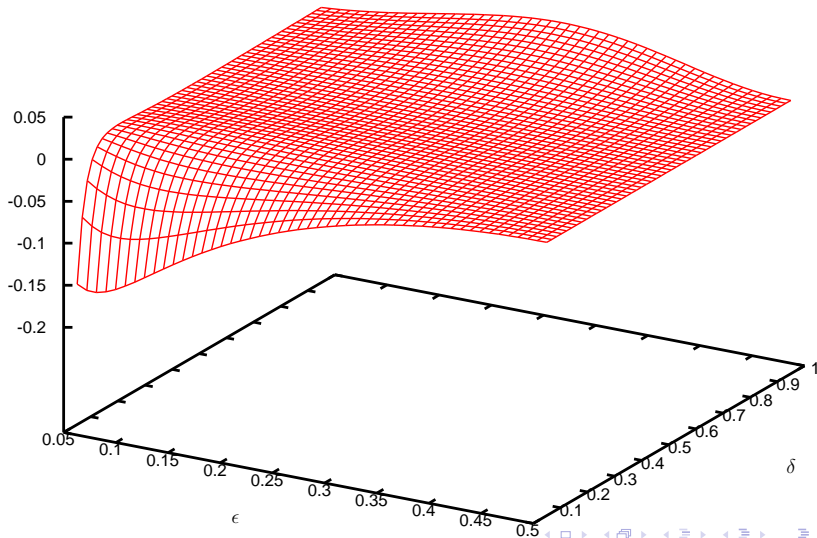
$$\begin{aligned} \mathcal{X} &= \{0, 1\} & p(x_1 = 0) &= 0.5 & p(x_n = x_{n-1}) &= 1 - \delta \\ \mathcal{Y} &= \mathcal{X} & & & p(y_n = x_n) &= 1 - \epsilon. \end{aligned}$$

- ▶ δ controls ergodicity of the state process.
- ▶ ϵ controls the information contained in observations.

Consider estimating $\mathbb{E}(X_2 | Y_{1:2} = (0, 1))$.

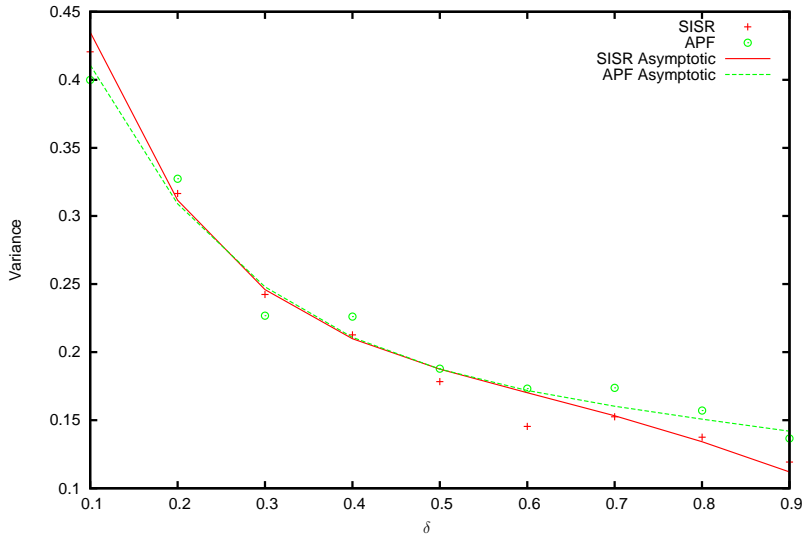


Variance of APF - Variance of SIR





Practical Implications

Variance Comparison at $\epsilon = 0.25$ 

○
○○○
○○

○○
○○
○○○○

○○○○○○○
○○
○○
○

○

Applications & Extensions



SMC Samplers (Del Moral, Doucet & Jasra, 2006)

SIR Techniques can be adapted for any sequence of distributions $\{\pi_n\}$.

- ▶ Define

$$\widetilde{\pi}_n(x_{1:n}) = \pi_n(x_n) \prod_{k=1}^{n-1} L_k(x_{k+1}, x_k).$$

- ▶ Sample at time n ,

$$X_n^i \sim K_n(X_{n-1}^i, \cdot)$$

- ▶ Weight

$$W_n^i \propto \frac{\pi_n(X_n^i) L_{n-1}(X_n^i, X_{n-1}^i)}{\pi_{n-1}(X_{n-1}^i) K_n(X_{n-1}^i, X_n^i)}$$

- ▶ Resample.



Auxiliary SMC Samplers

An *auxiliary SMC sampler* for a sequence of distributions π_n comprises:

- ▶ An SMC sampler targeting some auxiliary sequence of distributions μ_n .
- ▶ A sequence of importance weight functions

$$\widetilde{w}_n(x_n) \propto \frac{d\pi_n}{d\mu_n}(x_n).$$

Generic approaches:

- ▶ Choose $\mu_n(x_n) \propto \pi_n(x_n)V_n(x_n)$.
- ▶ Incorporate as much information as possible prior to resampling.



Auxiliary SMC Samplers

An *auxiliary SMC sampler* for a sequence of distributions π_n comprises:

- ▶ An SMC sampler targeting some auxiliary sequence of distributions μ_n .
- ▶ A sequence of importance weight functions

$$\widetilde{w}_n(x_n) \propto \frac{d\pi_n}{d\mu_n}(x_n).$$

Generic approaches:

- ▶ Choose $\mu_n(x_n) \propto \pi_n(x_n)V_n(x_n)$.
- ▶ Incorporate as much information as possible prior to resampling.



Resample-Move Approaches

A common approach in SMC Samplers:

- ▶ Choose $K_n(x_{n-1}, x_n)$ to be π_n -invariant.
- ▶ Set

$$L_{n-1}(x_n, x_{n-1}) = \frac{\pi_n(x_{n-1})K_n(x_{n-1}, x_n)}{\pi_n(x_n)}.$$

- ▶ Leading to

$$W_n(x_{n-1}, x_n) = \frac{\pi_n(x_{n-1})}{\pi_{n-1}(x_{n-1})}.$$

- ▶ It would make more sense to resample and then move...



An Interpretation of Pure Resample-Move

It's SIR with auxiliary distributions...

$$\mu_n(x_n) = \pi_{n+1}(x_n)$$

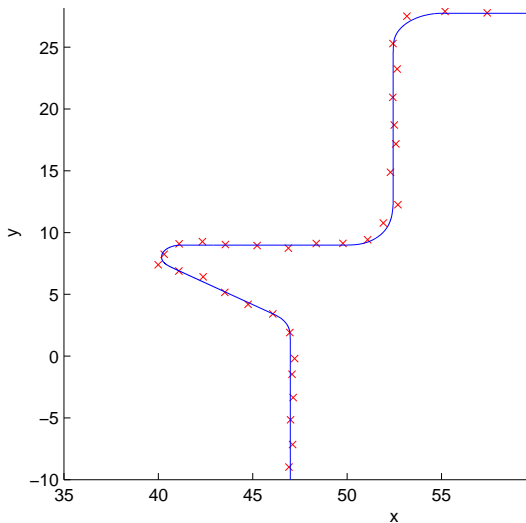
$$L_{n-1}(x_n, x_{n-1}) = \frac{\mu_{n-1}(x_{n-1})K_n(x_{n-1}, x_n)}{\mu_{n-1}(x_n)} = \frac{\pi_n(x_{n-1})K_n(x_{n-1}, x_n)}{\pi_n(x_n)}$$

$$w_n(x_{n-1:n}) = \frac{\mu_n(x_n)}{\mu_{n-1}(x_n)} = \frac{\pi_{n+1}(x_n)}{\pi_n(x_n)}$$

$$\tilde{w}_n(x_n) = \frac{\mu_{n-1}(x_n)}{\mu_n(x_n)} = \frac{\pi_n(x_n)}{\pi_{n+1}(x_n)}.$$



Piecewise-Deterministic Processes





“Filtering” of Piecewise-Deterministic Processes (Whiteley, Johansen & Godsill, 2007)

- ▶ $X_n = (k_n, \tau_{n,1:k_n}, \theta_{n,0:k_n})$ specifies a continuous time trajectory $(\zeta_t)_{t \in [0, t_n]}$.
- ▶ ζ_t observed in the presence of noise, e.g. $Y_n = \zeta_{t_n} + V_n$
- ▶ Target sequence of distributions $\{\pi_n\}$ on nested spaces:

$$\begin{aligned} \pi_n(k, \tau_{1:k}, \theta_{0:k}) &\propto q(\theta_0) \prod_{j=1}^k f(\tau_j | \tau_{j-1}) q(\theta_j | \tau_j, \theta_{j-1}, \tau_{j-1}) \\ &\times S(t_n, \tau_k) \prod_{p=1}^n g(y_p | \zeta_{t_p}) \end{aligned}$$



“Filtering” of Piecewise-Deterministic Processes (Whiteley, Johansen & Godsill, 2007)

- ▶ π_n yields posterior marginal distributions for recent history of ζ_t
- ▶ Employ auxiliary SMC samplers
- ▶ Choose $\mu_n(k, \tau_{1:k}, \theta_{0:k}) \propto \pi_n(k, \tau_{1:k}, \theta_{0:k}) V_n(\tau_k, \theta_k)$
- ▶ Kernel proposes new pairs (τ_j, θ_j) and adjusts recent pairs
- ▶ Applications in object-tracking and finance



Probability Hypothesis Density Filtering (Mahler, 2003; Vo et al. 2005)

- ▶ Approximates the optimal filter for a class of spatial point process-valued HMMs
- ▶ A recursion for intensity functions:

$$\alpha_n(x_n) = \int_E f(x_n|x_{n-1})p_S(x_{n-1})\hat{\alpha}_{n-1}(x_{n-1})dx_{n-1} + \gamma(x_n)$$

$$\hat{\alpha}_n(x_n) = \left[1 - p_D(x_n) + \sum_{r=1}^{m_n} \frac{\psi_{n,r}(x_n)}{\mathcal{Z}_{n,r}} \right] \alpha_n(x_n).$$

- ▶ Approximate the intensity functions $\{\hat{\alpha}_n\}_{n \geq 0}$ using SMC



An ASMC Implementation (Whiteley et al., 2007)

- ▶ $\hat{\alpha}_{n-1}(dx_{n-1}) \approx \hat{\alpha}_{n-1}^N(dx_{n-1}) = \frac{1}{N} \sum_{i=1}^N W_{n-1}^i \delta_{X_{n-1}^i}(dx_{n-1})$
- ▶ In a simple case, target integral at n th iteration:

$$\int_E \int_E \sum_{r=1}^{m_n} \varphi(x_n) \frac{\psi_{n,r}(x_n)}{\mathcal{Z}_{n,r}} f(x_n|x_{n-1}) p_S(x_{n-1}) dx_n \hat{\alpha}_{n-1}^N(dx_{n-1})$$

- ▶ Proposal distribution $q_n(x'_n, x'_{n-1}, r_n)$ built from $\hat{\alpha}_{n-1}^N$, potential functions $\{V_{n,r}\}$ and factorises:

$$q_n(x'_n|x'_{n-1}, r) = \frac{\sum_{i=1}^N V_{n,r}(X_{n-1}^{(i)}) W_{n-1} \delta_{X_{n-1}^{(i)}}(dx'_{n-1})}{\sum_{i=1}^N V_{n,r}(X_{n-1}^{(i)}) W_{n-1}} q_n(r)$$

- ▶ and also estimate normalising constants $\{\mathcal{Z}_{n,r}\}$ by IS



Stratification...

- ▶ Back to HMM, let $(A_p)_{p=1}^M$ denote a partition of \mathcal{X}
- ▶ Introduce stratum indicator variable $R_n = \sum_{p=1}^M p \mathbb{I}_{A_p}(X_n)$
- ▶ Define extended model:

$$\begin{aligned}
 p(x_{1:n}, r_{1:n} | y_{1:n}) &\propto g(y_n | x_n) f(x_n | r_n, x_{n-1}) f(r_n | x_{n-1}) \\
 &\quad \times p(x_{1:n-1}, r_{1:n-1} | y_{1:n-1})
 \end{aligned}$$

- ▶ An auxiliary SMC filter resamples from distribution with $N \times M$ support points, over *particles* \times *strata*
- ▶ Applications in switching state space models

○
○○○
○○

○○
○○
○○○○

○○○○○○○
○○
○○
○

○

Conclusions

○
○○○
○○

○○
○○
○○○○

○○○○○○○
○○
○○
○

●

Conclusions

- ▶ The APF is a standard SIR algorithm for nonstandard distributions.
- ▶ This interpretation...
 - ▶ allows standard results to be applied directly,
 - ▶ provides guidance on implementation,
 - ▶ and allows the same technique to be applied in more general settings.
- ▶ Thanks for listening... any questions?

○
○○○
○○○○
○○
○○○○○○○○○○○
○○
○

●

Conclusions

- ▶ The APF is a standard SIR algorithm for nonstandard distributions.
- ▶ This interpretation...
 - ▶ allows standard results to be applied directly,
 - ▶ provides guidance on implementation,
 - ▶ and allows the same technique to be applied in more general settings.
- ▶ Thanks for listening... any questions?

○
○○○
○○

○○
○○
○○○○

○○○○○○○
○○
○

●

Conclusions

- ▶ The APF is a standard SIR algorithm for nonstandard distributions.
- ▶ This interpretation...
 - ▶ allows standard results to be applied directly,
 - ▶ provides guidance on implementation,
 - ▶ and allows the same technique to be applied in more general settings.
- ▶ Thanks for listening... any questions?

○
○○○
○○

○○
○○
○○○○

○○○○○○○
○○
○

●

Conclusions

- ▶ The APF is a standard SIR algorithm for nonstandard distributions.
- ▶ This interpretation...
 - ▶ allows standard results to be applied directly,
 - ▶ provides guidance on implementation,
 - ▶ and allows the same technique to be applied in more general settings.
- ▶ Thanks for listening... any questions?

○
○○○
○○

○○
○○
○○○○

○○○○○○○
○○
○

●

Conclusions

- ▶ The APF is a standard SIR algorithm for nonstandard distributions.
- ▶ This interpretation...
 - ▶ allows standard results to be applied directly,
 - ▶ provides guidance on implementation,
 - ▶ and allows the same technique to be applied in more general settings.
- ▶ Thanks for listening... any questions?

○
○○○
○○○○
○○
○○○○○○○○○○○
○○
○

●

Conclusions

- ▶ The APF is a standard SIR algorithm for nonstandard distributions.
- ▶ This interpretation...
 - ▶ allows standard results to be applied directly,
 - ▶ provides guidance on implementation,
 - ▶ and allows the same technique to be applied in more general settings.
- ▶ Thanks for listening... any questions?

○
○○○
○○○○
○○
○○○○○○○○○○○
○○
○

●

Conclusions

- ▶ The APF is a standard SIR algorithm for nonstandard distributions.
- ▶ This interpretation...
 - ▶ allows standard results to be applied directly,
 - ▶ provides guidance on implementation,
 - ▶ and allows the same technique to be applied in more general settings.
- ▶ Thanks for listening... any questions?

References

- P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society B*, 63(3):411–436, 2006.
- A. M. Johansen and A. Doucet. Auxiliary variable sequential Monte Carlo methods. Technical Report 07:09, University of Bristol, Department of Mathematics – Statistics Group, University Walk, Bristol, BS8 1TW, UK, July 2007.
- A. M. Johansen and A. Doucet. A note on the auxiliary particle filter. *Statistics and Probability Letters*, 2008. To appear.
- A. M. Johansen and N. Whiteley. A modern perspective on auxiliary particle filters. In *Proceedings of Workshop on Inference and Estimation in Probabilistic Time Series Models*. Isaac Newton Institute, June 2008. To appear.
- R. P. S. Mahler. Multitarget Bayes filtering via first-order multitarget moments. *IEEE Transactions on Aerospace and Electronic Systems*, 39(4):1152, October 2003.
- M. K. Pitt and N. Shephard. Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, 94(446):590–599, 1999.
- B. Vo, S.S. Singh, and A. Doucet. Sequential Monte Carlo methods for multi-target filtering with random finite sets. *IEEE Transactions on Aerospace and Electronic Systems*, 41(4):1223–1245, October 2005.
- N. Whiteley, A. M. Johansen, and S. Godsill. Monte Carlo filtering of piecewise-deterministic processes. Technical Report CUED/F-INFENG/TR-592, University of Cambridge, Department of Engineering, Cambridge University Engineering Department, Trumpington Street, Cambridge, CB2 1PZ, December 2007.
- N. Whiteley, S. Singh, and S. Godsill. Auxiliary particle implementation of the probability hypothesis density filter. Technical Report CUED F-INFENG/590, University of Cambridge, Department of Engineering, Trumpington Street, Cambridge, CB1 2PZ, United Kingdom, 2007.