Sequential Monte Carlo:

Selected Methodological Applications

Adam M. Johansen

a.m.johansen@warwick.ac.uk

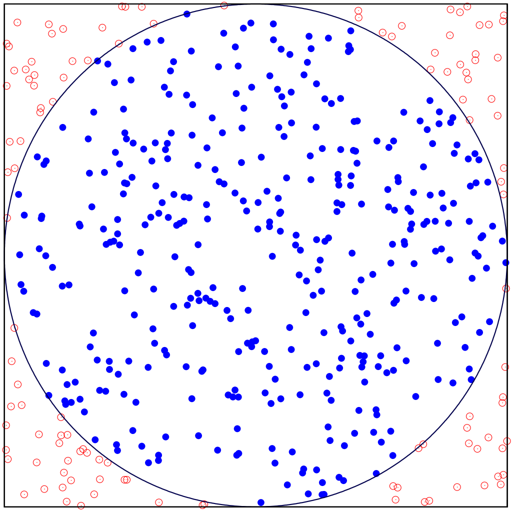
Warwick University Centre for Scientific Computing

Outline

- Sequential Monte Carlo
- Applications
  - Parameter Estimation
  - Rare Event Simulation
  - Filtering of Piecewise Deterministic Processes

Background

Estimating π



- Rain is uniform.
- Circle is inscribed in square.
- $A_{\text{square}} = 4r^2$.
- $A_{\text{circle}} = \pi r^2$.
- $p = \frac{A_{\text{circle}}}{A_{\text{square}}} = \frac{\pi}{4}$.
- 383 of 500 "successes".
- $\hat{\pi} = 4\frac{383}{500} = 3.06$.
- Also obtain confidence intervals.

The Monte Carlo Method

- Given a probability density, f,

$$I = \int_E \varphi(x) f(x) dx$$

- Simple Monte Carlo solution:
  - Sample $X_1, \ldots, X_N \overset{iid}{\sim} f$.
  - Estimate $\hat{I} = \frac{1}{N} \sum_{i=1}^{n} \varphi(X_N)$.
- Justified by the law of large numbers…
- and the central limit theorem.

Importance Sampling

- Given g, such that
  - $f(x) > 0 \Rightarrow g(x) > 0$
  - and $f(x)/g(x) < \infty$,

  define $w(x) = f(x)/g(x)$ and:

$$I = \int \varphi(x) f(x) dx = \int \varphi(x) w(x) g(x) dx.$$

- This suggests the importance sampling estimator:
  - Sample $X_1, \ldots, X_N \overset{iid}{\sim} g$.
  - Estimate $\hat{I} = \frac{1}{N} \sum_{i=1}^{N} w(X_i) \varphi(X_i)$.

Markov Chain Monte Carlo

- Typically difficult to construct a good proposal density.
- MCMC works by constructing an ergodic Markov chain of invariant distribution π, X_n using it's ergodic averages:

$$\frac{1}{N}\sum_{i=1}^{N}\varphi(X_i)$$

  to approach $\mathbb{E}_\pi[\varphi]$.
- Justified by ergodic theorems / central limit theorems.
- We aren't going to take this approach.

A Motivating Example: Filtering

- Let $X_1, \ldots$ denote the position of an object which follows Markovian dynamics.
- Let $Y_1, \ldots$ denote a collection of observations: $Y_i | X_i = x_i \sim g(\cdot | x_i)$.
- We wish to estimate, as observations arrive, $p(x_{1:n} | y_{1:n})$.
- A recursion obtained from Bayes rule exists but is intractable in most cases.

More Generally

- The problem in the previous example is really tracking a sequence of distributions.
- Key structural property of the smoothing distributions: increasing state spaces.
- Other problems with the same structure exist.
- Any problem of sequentially approximating a sequence of such distributions, p_n, can be addressed in the same way.

Importance Sampling in This Setting

- Given $p_n(x_{1:n})$ for $n = 1, 2, \ldots$.
- We could sample from a sequence $q_n(x_{1:n})$ for each n.
- Or we could let $q_n(x_{1:n}) = q_n(x_n|x_{1:n-1})q_{n-1}(x_{1:n})$ and re-use our samples.
- The importance weights become:

$$
\begin{aligned}
w_n(x_{1:n}) \propto \frac{p_n(x_{1:n})}{q_n(x_{1:n})} &= \frac{p_n(x_{1:n})}{q_n(x_n|x_{1:n-1})q_{n-1}(x_{1:n-1})} \\
&= \frac{p_n(x_{1:n})}{q_n(x_n|x_{1:n-1})p_{n-1}(x_{1:n-1})} w_{n-1}(x_{1:n-1})
\end{aligned}
$$

Sequential Importance Sampling

At time 1.

For $i = 1 : N$, sample $X_1^{(i)} \sim q_1(\cdot)$.

For $i = 1 : N$, compute $W_1^i \propto w_1\left(X_1^{(i)}\right) = \frac{p_1\left(X_1^{(i)}\right)}{q_1\left(X_1^{(i)}\right)}$.

At time n, $n \geq 2$.

Sampling Step

For $i = 1 : N$, sample $X_n^{(i)} \sim q_n\left(\cdot \middle| X_{n-1}^{(i)}\right)$.

Weighting Step

For $i = 1 : N$, compute

$$w_n\left(X_{1:n-1}^{(i)}, X_n^{(i)}\right) = \frac{p_n\left(X_{1:n-1}^{(i)}, X_n^{(i)}\right)}{p_{n-1}\left(X_{1:n-1}^{(i)}\right) q_n\left(X_n^{(i)} \middle| X_{n-1}^{(i)}\right)}$$

and $W_n^{(i)} \propto W_{n-1}^{(i)} w_n\left(X_{1:n-1}^{(i)}, X_n^{(i)}\right)$.

Sequential Importance Resampling

At time n, $n \geq 2$.

Sampling Step

For $i = 1 : N$, sample $X_{n,n}^{(i)} \sim q_n\left(\cdot \middle| \widetilde{X}_{n-1}^{(i)}\right)$.

Resampling Step

For $i = 1 : N$, compute

$$w_n\left(\widetilde{X}_{n-1}^{(i)}, X_{n,n}^{(i)}\right) = \frac{p_n\left(\widetilde{X}_{n-1}^{(i)}, X_{n,n}^{(i)}\right)}{p_{n-1}\left(\widetilde{X}_{n-1}^{(i)}\right) q_n\left(X_{n,n}^{(i)} \middle| \widetilde{X}_{n-1}^{(i)}\right)}$$

and $W_n^{(i)} = \frac{w_n\left(\widetilde{X}_{n-1}^{(i)}, X_{n,n}^{(i)}\right)}{\sum_{j=1}^{N} w_n\left(\widetilde{X}_{n-1}^{(j)}, X_{n,n}^{(j)}\right)}$.

For $i = 1 : N$, sample $\widetilde{X}_n^{(i)} \sim \sum_{j=1}^{N} W_n^{(j)} \delta_{\left(\widetilde{X}_{n-1}^{(j)}, X_{n,n}^{(j)}\right)}(dx_{1:n})$.

SMC Samplers

Actually, these techniques can be used to sample from *any* sequence of distributions (Del Moral et al., 2006).

- Given a sequence of *target* distributions, η_n, on $E_n \ldots$,
- construct a synthetic sequence $\widetilde{\eta}_n$ on spaces $\bigotimes_{p=1}^{n} E_p$
- by introducing Markov kernels, L_p from E_{p+1} to E_p:

$$\widetilde{\eta}_n(x_{1:n}) = \eta_n(x_n) \prod_{p=1}^{n-1} L_p\left(x_{p+1}, x_p\right),$$

- These distributions
  - have the target distributions as time marginals,
  - have the correct structure to employ SMC techniques.

SMC Outline

- Given a sample $\{X_{1:n-1}^{(i)}\}_{i=1}^{N}$ targeting $\widetilde{\eta}_{n-1}$,
- sample $X_n^{(i)} \sim K_n(X_{n-1}^{(i)}, \cdot)$,
- calculate

$$W_n(X_{1:n}^{(i)}) = \frac{\eta_n(X_n^{(i)}) L_{n-1}(X_n^{(i)}, X_{n-1}^{(i)})}{\eta_{n-1}(X_{n-1}^{(i)}) K_n(X_{n-1}^{(i)}, X_n^{(i)})}.$$

- Resample, yielding: $\{X_{1:n}^{(i)}\}_{i=1}^{N}$ targeting $\widetilde{\eta}_n$.
- Hints that we'd like to use

$$L_{n-1}(x_n, x_{n-1}) = \frac{\eta_{n-1}(x_{n-1}) K_n(x_{n-1}, x_n)}{\int \eta_{n-1}(x'_{n-1}) K_n(x'_{n-1}, x_n)}.$$

Parameter Estimation in Latent Variable Models

Joint work with Arnaud Doucet and Manuel Davy.

Maximum {Likelihood|*a Posteriori*} Estimation

- Consider a model with:
  - parameters, θ,
  - latent variables, x, and
  - observed data, y.
- Aim to maximise Marginal likelihood

$$p(y|\theta) = \int p(x, y|\theta) dx$$

  or posterior

$$p(\theta|y) \propto \int p(x, y|\theta) p(\theta) dx.$$

- Traditional approach is Expectation-Maximisation (EM)
  - Requires objective function in closed form.
  - Susceptible to trapping in local optima.

A Probabilistic Approach

- A distribution of the form

$$\pi(\theta|y) \propto p(\theta)p(y|\theta)^{\gamma}$$

  will become concentrated, as $\gamma \to \infty$ on the maximisers of $p(y|\theta)$ under weak conditions (Hwang, 1980).
- **Key point:** Synthetic distributions of the form:

$$\bar{\pi}_{\gamma}(\theta, x_{1:\gamma}|y) \propto p(\theta) \prod_{i=1}^{\gamma} p(x_i, y|\theta)$$

  admit the marginals

$$\bar{\pi}_{\gamma}(\theta|y) \propto p(\theta)p(y|\theta)^{\gamma}.$$

Maximum Likelihood via SMC

- Use a sequence of distributions $\eta_n = \pi_{\gamma_n}$ for some $\{\gamma_n\}$.
- Has previously been suggested in an MCMC context (Doucet et al., 2002).
  - Requires extremely slow "annealing".
  - Separation between distributions is large.
- SMC has two main advantages:
  - Introducing bridging distributions, for $\gamma = \lfloor \gamma \rfloor + \langle \gamma \rangle$, of:

$$\bar{\pi}_\gamma(\theta, x_{1:\lfloor \gamma \rfloor + 1} | y) \propto p(\theta) p(x_{\lfloor \gamma \rfloor + 1}, y | \theta)^{\langle \gamma \rangle} \prod_{i=1}^{\lfloor \gamma \rfloor} p(x_i, y | \theta)$$

    is straightforward.
  - Population of samples improves robustness.

Three Algorithms

- A generic SMC sampler can be written down directly…
- Easy case:
  - Sample from $p(x_n|y, \theta_{n-1})$ and $p(\theta_n|x_n, y)$.
  - Weight according to $p(y|\theta_{n-1})^{\gamma_n - \gamma_{n-1}}$.
- General case:
  - Sample existing variables from a η_{n-1}-invariant kernel:

$$(\theta_n, X_{n,1:\gamma_{n-1}}) \sim \mathcal{K}_{n-1}((\theta_{n-1}, X_{n-1}), \cdot).$$

  - Sample new variables from an arbitrary proposal:

$$X_{n,\gamma_{n-1}+1:\gamma_n} \sim q(\cdot|\theta_n).$$

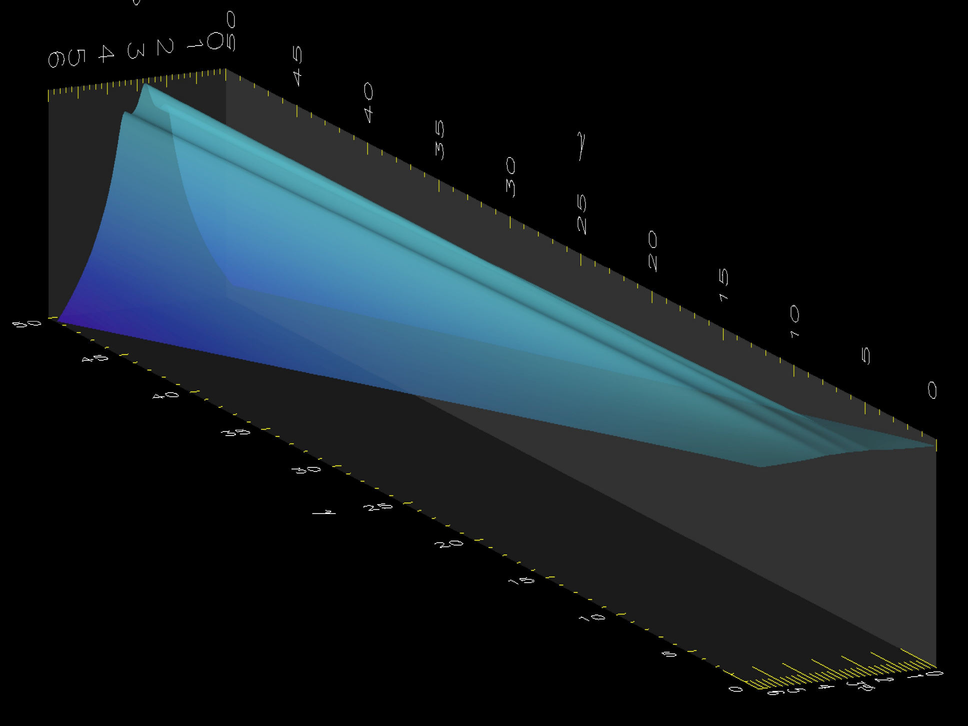
  - Use the composition of a time-reversal and optimal auxiliary kernel.
  - Weight expression does not involve the marginal likelihood.

Toy Example

- Student t-distribution of unknown location parameter θ with $\nu = 0.05$.
- Four observations are available, $y = (-20, 1, 2, 3)$.
- Log likelihood is:

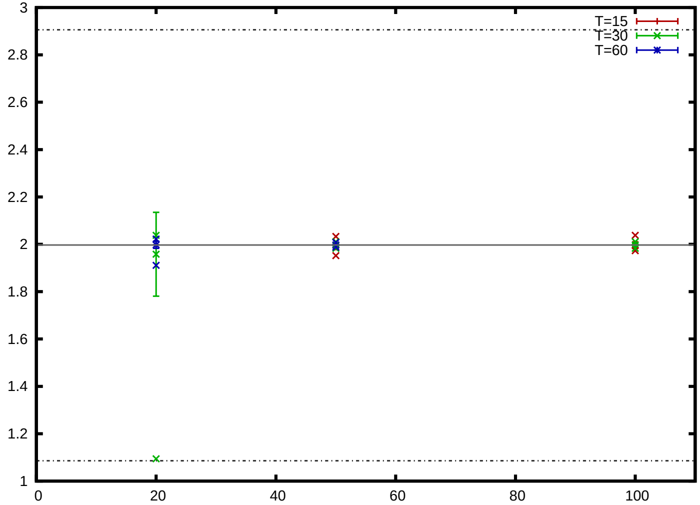
$$\log p(y|\theta) = -0.525 \sum_{i=1}^{4} \log\left(0.05 + (y_i - \theta)^2\right).$$

- Global maximum is at 1.997.
- Local maxima at $\{-19.993, 1.086, 2.906\}$.



y
t
0
5
10
15
20
25
30
35
40
45
50

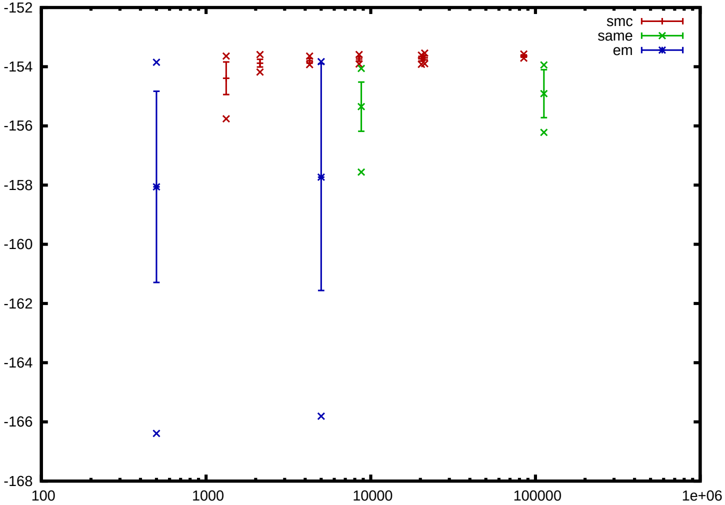
It actually works. . .



Example: Gaussian Mixture Model – MAP Estimation

- Likelihood $p(y|x, \omega, \mu, \sigma) = \mathcal{N}(y|\mu_x, \sigma_x^2)$.
- Marginal likelihood $p(y|\omega, \mu, \sigma) = \sum_{j=1}^{3} \omega_j \mathcal{N}(y|\mu_j, \sigma_j^2)$.
- Diffuse conjugate priors were employed.
- All full conditional distributions of interest are available.
- Marginal posterior can be calculated.

Example: GMM (Galaxy Data Set)



Rare Event Simulation

Joint work with Pierre Del Moral and Arnaud Doucet.

Problem Formulation

- Consider the canonical Markov chain:

$$\left(\Omega = \prod_{t=0}^{\infty} E_t, \mathcal{F} = \prod_{t=0}^{\infty} \mathcal{F}_t, (X_t)_{t\in\mathbb{N}}, \mathbb{P}_{\mu_0}\right),$$

- The law $\mathbb{P}_{\mu_0}$ is defined by its finite dimensional distributions:

$$\mathbb{P}_{\mu_0} \circ X_{0:p}^{-1}(dx_{0:p}) = \mu_0(dx_0) \prod_{i=1}^{p} M_i(x_{i-1}, dx_i).$$

- We are interested in *rare events*.

Static Rare Events

We term the first type of rare events which we consider *static rare events*:

- The first $P+1$ elements of the canonical Markov chain lie in a rare set, $\mathcal{T}$.
- That is, we are interested in

$$\mathbb{P}_{\mu_0}\left(x_{0:P} \in \mathcal{T}\right)$$

and

$$\mathbb{P}_{\mu_0}\left(x_{0:P} \in dx_{0:P} \left| x_{0:P} \in \mathcal{T}\right.\right)$$

- We assume that the rare event is characterised as a level set of a suitable potential function:

$$V : \mathcal{T} \to [\hat{V}, \infty), \text{ and } V : E_{0:P} \setminus \mathcal{T} \to (-\infty, \hat{V}).$$

Dynamic Rare Events

The other class of rare events in which we are interested are termed *dynamic rare events*:

- A Markov process hits some rare set, $\mathcal{T}$, before its first entrance to some recurrent set $\mathcal{R}$.
- That is, given the stopping time $\tau = \inf\{p : X_p \in \mathcal{T} \cup \mathcal{R}\}$, we seek

$$\mathbb{P}_{\mu_0}\left(X_\tau \in \mathcal{T}\right)$$

and the associated conditional distribution:

$$\mathbb{P}_{\mu_0}\left(\tau = t, X_{0:t} \in dx_{0:t} | X_\tau \in \mathcal{T}\right)$$

Intuition

- Principle novelty: applying an efficient sampling technique which allows us to operate directly on the path space of the Markov chain.
- Two components to this approach:
  - Constructing a sequence of synthetic distributions
  - Applying sequential importance sampling and resampling strategies.

Static Rare Events: Our Approach

- Initialise by sampling from the law of the Markov chain.
- Iteratively obtain samples from a sequence of distributions which moves "smoothly" towards the target.
- Proposed sequence of distributions:

$$\eta_n(dx_{0:P}) \propto \mathbb{P}_{\mu_0}(dx_{0:P}) g_{n/T}(x_{0:P})$$

$$g_\theta(x_{0:P}) = \left(1 + \exp\left(-\alpha(\theta)\left(V(x_{0:P}) - \hat{V}\right)\right)\right)^{-1}$$

- Estimate the normalising constant of the final distribution and correct via importance sampling.

Path Sampling [See ⋆⋆ or Gelman and Meng, 1998]

- Given a sequence of densities $p(x|\theta) = q(x|\theta)/z(\theta)$:

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \log z(\theta) = \mathbb{E}_\theta \left[\frac{\mathrm{d}}{\mathrm{d}\theta} \log q(\cdot|\theta) \right] \qquad (\star)$$

  where the expectation is taken with respect to $p(\cdot|\theta)$.
- Consequently, we obtain:

$$\log \left(\frac{z(1)}{z(0)} \right) = \int_0^1 \mathbb{E}_\theta \left[\frac{\mathrm{d}}{\mathrm{d}\theta} \log q(\cdot|\theta) \right]$$

- In our case, we use our particle system to approximate *both* integrals.

Approximate the path sampling identity to estimate the normalising constant:

$$\hat{Z}_1 = \frac{1}{2}\exp\left[\sum_{n=1}^{T}\left(\alpha(n/T)-\alpha((n-1)/T)\right)\frac{\hat{E}_{n-1}+\hat{E}_n}{2}\right]$$

$$\hat{E}_n = \frac{\sum_{j=1}^{N} W_n^{(j)} \frac{V\left(X_n^{(j)}\right)-\hat{V}}{1+\exp\left(\alpha_n\left(V\left(X_n^{(j)}\right)-\hat{V}\right)\right)}}{\sum_{j=1}^{N} W_n^{(j)}}$$

Estimate the rare event probability:

$$p^{\star} = \hat{Z}_1 \frac{\sum_{j=1}^{N} W_T^{(j)}\left(1+\exp(\alpha(1)(V\left(X_T^{(j)}\right)-\hat{V}))\right)\mathbb{I}_{(\hat{V},\infty]}\left(V\left(X_T^{(j)}\right)\right)}{\sum_{j=1}^{N} W_T^{(j)}}.$$

Example: Gaussian Random Walk

- A toy example: $M_t(R_{t-1}, R_t) = \mathcal{N}(R_t | R_{t-1}, 1)$.
- $\mathcal{T} = \mathbb{R}^P \times [\hat{V}, \infty)$.
- Proposal kernel:

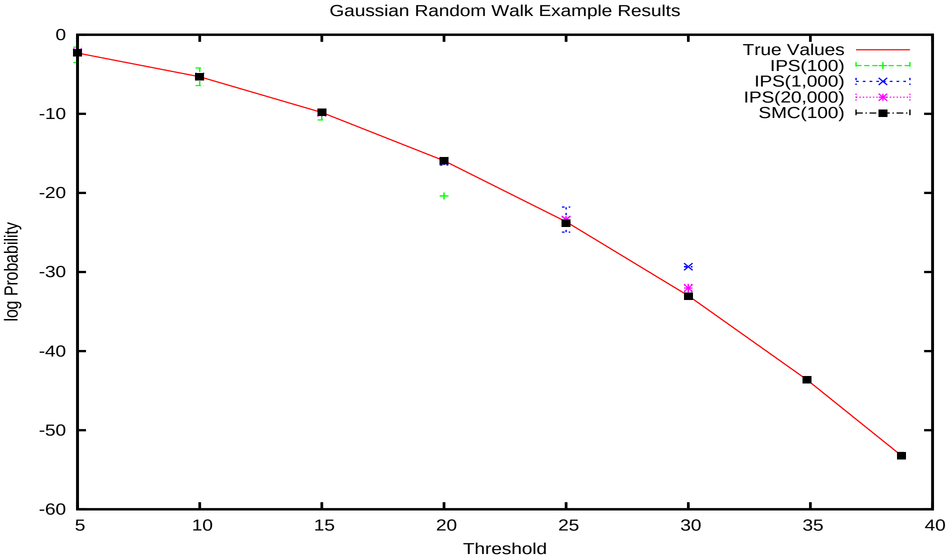
$$K_n(X_{n-1}, X_n) = \sum_{j=-S}^{S} \alpha_{n+1}(X_{n-1}, X_n) \prod_{i=1}^{P} \delta_{X_{n-1,i}+ij\delta}(X_{n,i}),$$

where the weighting of individual moves is given by

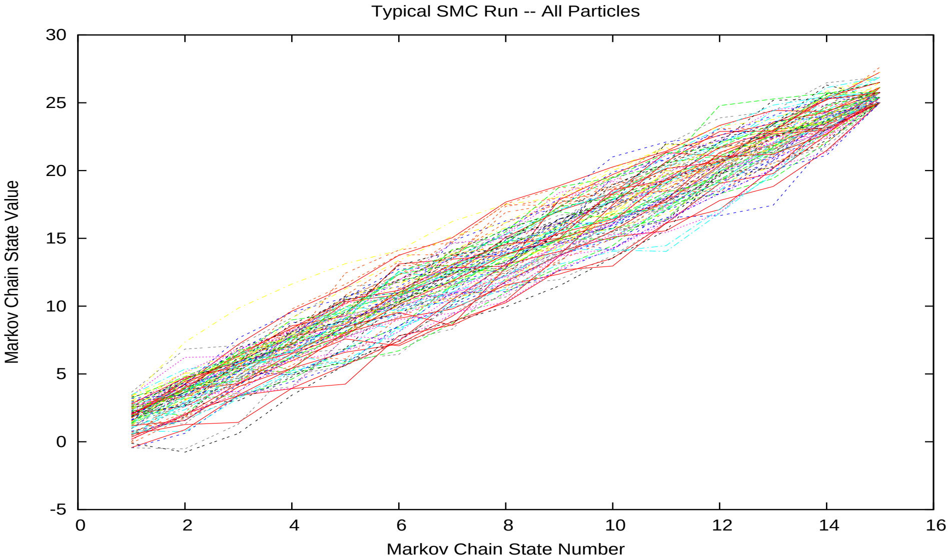
$$\alpha_n(X_{n-1}, X_n) \propto \eta_n(X_n).$$

- Linear annealing schedule.
- Number of distributions $T \propto \hat{V}^{3/2}$ (T=2500 when $\hat{V} = 25$).

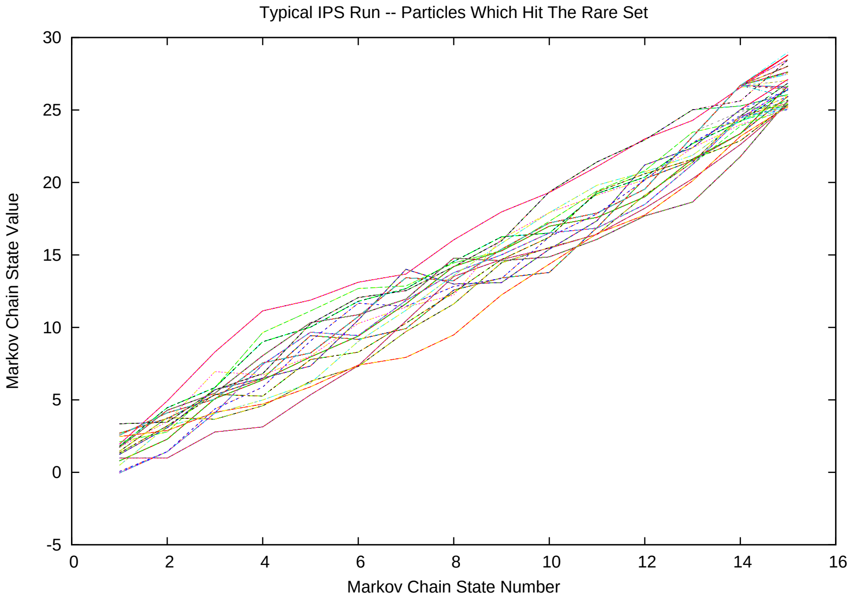
Example: Gaussian Random Walk



Example: Gaussian Random Walk



Example: Gaussian Random Walk



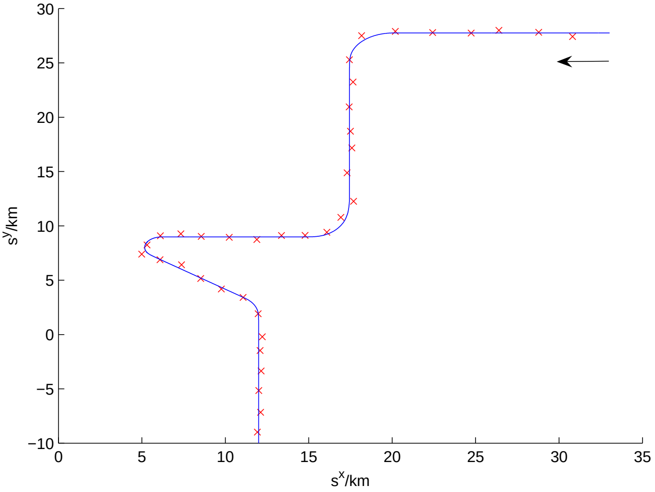
Filtering of Piecewise Deterministic Processes

Joint work with Nick Whiteley and Simon Godsill.

Motivation: Observing a Manoeuvering Object

- For $t \in \mathbb{R}_0^+$, consider object with position s_t, velocity v_t and acceleration a_t
- Summarise state by $\zeta_t = (s_t, v_t, a_t)$
- From initial condition ζ_0, state evolves until random time τ_1, at which acceleration jumps to a new random value, yielding ζ_{τ_1}
- From ζ_{τ_1}, evolution until τ_2, state becomes ζ_{τ_2}, etc.
- Observation times, $(t_n)_{n \in \mathbb{N}}$, at each t_n a noisy measurement of the object's position is made

Filtering of PD Processes



An Abstract Formulation

- Pair Markov chain $(\tau_j, \theta_j)_{j\in\mathbb{N}}$, $\tau_j \in \mathbb{R}^+$, $\theta_j \in \Theta$

$$p(d(\tau_j, \theta_j)|\tau_{j-1}, \theta_{j-1}) = q(d\theta_j|\theta_{j-1}, \tau_j, \tau_{j-1})f(d\tau_j|\tau_{j-1}),$$

- Count the jumps $\nu_t := \sum_j \mathbb{I}_{[\tau_j \leq t]}$
- Deterministic evolution function $F : \mathbb{R}_0^+ \times \Theta \to \Theta$, s.t. $\forall \theta \in \Theta$,

$$F(0, \theta) = \theta$$

- Signal process $(\zeta_t)_{t\in\mathbb{R}_0^+}$,

$$\zeta_t := F(t - \tau_{\nu_t}, \theta_{\nu_t})$$

Filtering 1

- This describes a Piecewise Deterministic Process.
- It's partially observed via observations $(Y_n)_{n\in\mathbb{N}}$, e.g.,

$$Y_n = G(\zeta_{t_n}) + V_n$$

  and likelihood function $g_n(y_n|\zeta_{t_n})$
- Filtering: given observations, $y_{1:n}$, estimate ζ_{t_n}.
- How can we approximate $p(\zeta_{t_n}|y_{1:n})$, $p(\zeta_{t_{n+1}}|y_{1:n+1})$, ... ?

Filtering 2

- Sequence of spaces $(E_n)_{n\in\mathbb{N}}$,

$$E_n = \biguplus_{k=0}^{\infty} \{k\} \times \mathbb{T}_{n,k} \times \Theta^{k+1},$$

$$\mathbb{T}_{n,k} = \{\tau_{1:k} : 0 < \tau_1 < \tau_2 < \ldots < \tau_k \leq t_n\}.$$

- Define $k_n := \nu_{t_n}$ and $X_n = (\zeta_0, k_n, \tau_{1:k_n}, \theta_{1:k_n}) \in E_n$
- Sequence of posterior distributions $(\eta_n)_{n\in\mathbb{N}}$

$$\eta_n(x_n) \propto q(\zeta_0) \prod_{j=1}^{k_n} f(\tau_j|\tau_{j-1}) q(\theta_j|\theta_{j-1}, \tau_j, \tau_{j-1}) \times \prod_{p=1}^{n} g_p(y_p|\zeta_{t_p}) S(\tau_{k_n}, t_n)$$

SMC Filtering

- Recall $X_n = (\zeta_0, k_n, \tau_{1:k_n}, \theta_{1:k_n})$ specifies a path $(\zeta_t)_{t\in[0,t_n]}$
- If forward kernel K_n only alters the recent components of x_{n-1} and adds new jumps/parameters in $E_n \setminus E_{n-1}$, online operation is possible

$$p(d\zeta_{t_n}|y_{1:n}) \approx \sum_{i=1}^{N} W_n^{(i)} \delta_{F(t_n-\tau_{k_n}^{(i)},\theta_{k_n}^{(i)})}(d\zeta_{t_n})$$

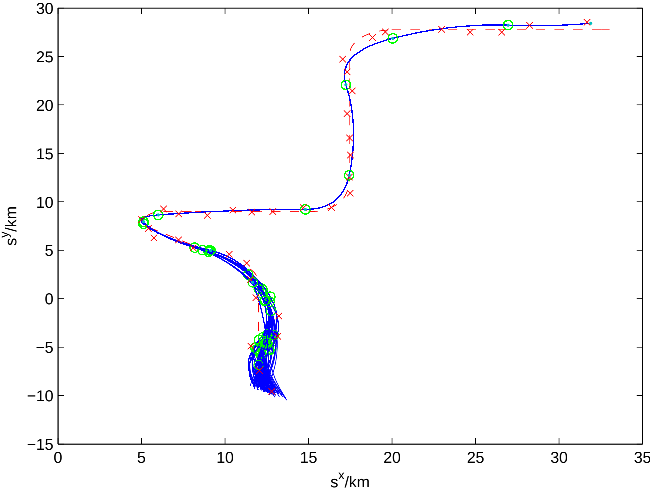
- A mixture proposal

$$K_n(x_{n-1}, x_n) = \sum_m \alpha_{n,m}(x_{n-1}) K_{n,m}(x_{n-1}, x_n),$$

SMC Filtering

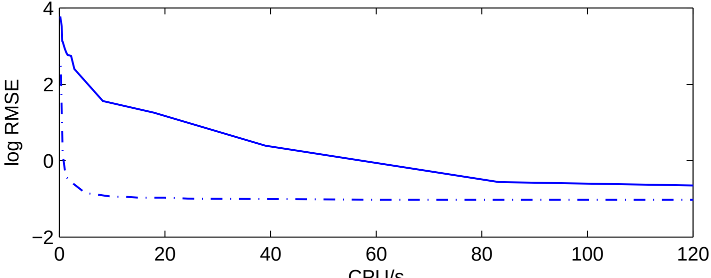
- When K_n corresponds to extending x_{n-1} into E_n by sampling from the prior, obtain the algorithm of (Godsill et al., 2007).
- This is inefficient as involves propagating multiple copies of particles after resampling
- A more efficient strategy is to propose births and to perturb the most recent jump time/parameter, (τ_k, θ_k)
- To minimize the variance the importance weights, we would like to draw from $\eta_n(\tau_k, \theta_k | x_{n-1} \setminus (\tau_k, \theta_k))$, or sensible approximations thereof.

Filtering of PD Processes



N	Godsill et al. 2007 RMSE / km	Godsill et al. 2007 CPU / s	Whiteley et al. 2007 RMSE / km	Whiteley et al. 2007 CPU / s
50	42.62	0.24	0.88	1.32
100	33.49	0.49	0.66	2.62
250	22.89	1.23	0.54	6.56
500	17.26	2.42	0.51	12.98
1000	12.68	5.00	0.50	26.07
2500	6.18	13.20	0.49	67.32
5000	3.52	28.79	0.48	142.84

Root mean square filtering error and CPU time, over 200 runs.



Convergence

- This framework allows us to analyse algorithm of Godsill et al. 2007
- $\mu_n(\varphi) := \int \varphi(\zeta_{t_n}) p(d\zeta_{t_n} | y_{1:n})$ and $\mu_n^N(\varphi)$ the corresponding SMC approximation
- Under standard regularity conditions

$$\sqrt{N}(\mu_n^N(\varphi) - \mu_n(\varphi)) \Rightarrow \mathcal{N}(0, \sigma_n^2(\varphi))$$

- Under rather strong assumptions*

$$\mathbb{E}\left[|\mu_n^N(\varphi) - \mu_n(\varphi)|^p\right]^{1/p} \leq \frac{c_p(\varphi)}{\sqrt{N}}$$

*which include: $(\zeta_{t_n})_{n \in \mathbb{N}}$ is uniformly ergodic Markov, likelihood bounded above and away from zero uniformly in time

Summary

SMCTC: C++ Template Class for SMC Algorithms

- Implementing SMC algorithms in C/C++ isn't hard.

- Software for implementing general SMC algorithms.
- C++ element largely confined to the library.
- Available (under a GPL-3 license from)

  `www2.warwick.ac.uk/fac/sci/statistics/staff/academic/johansen/smctc/`

  or type "smctc" into google.
- Example code includes estimation of Gaussian tail probabilities using the method described here.
- Particle filters can also be implemented easily.

In Conclusion

- Monte Carlo Methods have uses beyond the calculation of posterior means.
- SMC provides a viable alternative to MCMC.
- SMC is effective at:
  - ML and MAP estimation;
  - rare event estimation;
  - filtering outside the standard particle filtering framework.
  - ...
  - Other published applications include: approximate Bayesian computation, Bayesian estimation in GLMMs, options pricing and estimation in partially observed marked point processes.
- A huge amount of work remains to be done...

References

[1] P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo methods for Bayesian Computation. In *Bayesian Statistics 8*. Oxford University Press, 2006.

[2] P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society B*, 63(3):411–436, 2006.

[3] A. Doucet, S. J. Godsill, and C. P. Robert. Marginal maximum a posteriori estimation using Markov chain Monte Carlo. *Statistics and Computing*, 12:77–84, 2002.

[4] A. Gelman and X.-L. Meng. Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical Science*, 13(2):163–185, 1998.

[5] S. J. Godsill, J. Vermaak, K.-F. Ng, and J.-F. Li. Models and algorithms for tracking of manoeuvring objects using variable rate particle filters. *Proceedings of IEEE*, 95(5): 925–952, 2007.

[6] C.-R. Hwang. Laplace's method revisited: Weak convergence of probability measures. *The Annals of Probability*, 8(6):1177–1182, December 1980.

[7] A. M. Johansen. SMCTC: Sequential Monte Carlo in C++. Research Report 08:16, University of Bristol, Department of Mathematics – Statistics Group, University Walk, Bristol, BS8 1TW, UK, July 2008.

[8] A. M. Johansen, P. Del Moral, and A. Doucet. Sequential Monte Carlo samplers for rare events. In *Proceedings of the 6th International Workshop on Rare Event Simulation*, pages 256–267, Bamberg, Germany, October 2006.

[9] A. M. Johansen, A. Doucet, and M. Davy. Particle methods for maximum likelihood parameter estimation in latent variable models. *Statistics and Computing*, 18(1):47–57, March 2008.

[10] N. Whiteley, A. M. Johansen, and S. Godsill. Efficient Monte Carlo filtering for discretely observed jumping processes. In *Proceedings of IEEE Statistical Signal Processing Workshop*, pages 89–93, Madison, WI, USA, August 26th–29th 2007. IEEE.

[11] N. Whiteley, A. M. Johansen, and S. Godsill. Monte Carlo filtering of piecewise-deterministic processes. *In revision.*, 2008.

Path Sampling Identity

Given a probability density, $p(x|\theta) = q(x|\theta)/z(\theta)$:

$$\begin{aligned}
\frac{\partial}{\partial\theta}\log z(\theta) =& \frac{1}{z(\theta)}\frac{\partial}{\partial\theta}z(\theta) \\
=& \frac{1}{z(\theta)}\frac{\partial}{\partial\theta}\int q(x|\theta)dx \\
=& \int \frac{1}{z(\theta)}\frac{\partial}{\partial\theta}q(x|\theta)dx \qquad (\star\star) \\
=& \int \frac{p(x|\theta)}{q(x|\theta)}\frac{\partial}{\partial\theta}q(x|\theta)dx \\
=& \int p(x|\theta)\frac{\partial}{\partial\theta}\log q(x|\theta)dx = \mathbb{E}_{p(\cdot|\theta)}\left[\frac{\partial}{\partial\theta}\log q(x|\theta)\right]
\end{aligned}$$

wherever $\star\star$ is permissible. Back to $\star$.