Particle Filters

Inference via Interacting Particle Systems

Adam M. Johansen

a.m.johansen@warwick.ac.uk
http://www2.warwick.ac.uk/fac/sci/statistics/staff/academic/johansen/talks/

MASDOC Statistical Frontiers Seminar
13th January 2011

Mathematics and Statistics Centre for Doctoral Training

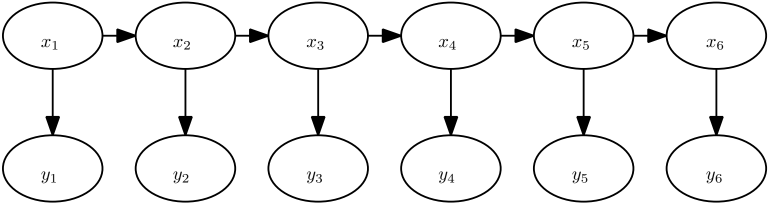
THE UNIVERSITY OF WARWICK

Part 1 – Statistics & Computation

Problems and Algorithms

A Statistical Problem

Hidden Markov Models / State Space Models



- Unobserved Markov chain $\{X_n\}$ transition f.
- Observed process $\{Y_n\}$ conditional density g.
- Density:

$$p(x_{1:n}, y_{1:n}) = f_1(x_1)g(y_1|x_1)\prod_{i=2}^{n} f(x_i|x_{i-1})g(y_i|x_i).$$

MASDOC

Filtering / Smoothing

- Let $X_1, \ldots$ denote the position of an object which follows Markovian dynamics:

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

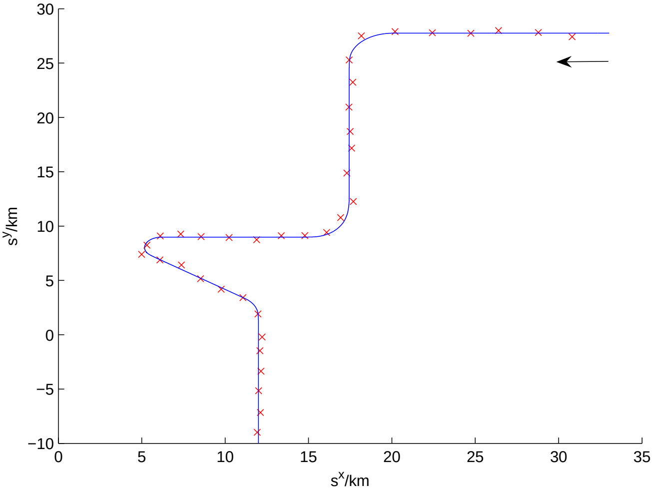
- Let $Y_1, \ldots$ denote a collection of observations:

$$Y_i | \{X_i = x_i\} \sim g(\cdot | x_i).$$

- Smoothing: estimate, as observations arrive, $p(x_{1:n} | y_{1:n})$.
- Filtering: estimate, as observations arrive, $p(x_n | y_{1:n})$.
- Formal Solution:

$$p(x_{1:n} | y_{1:n}) = p(x_{1:n-1} | y_{1:n-1}) \frac{f(x_n | x_{n-1}) g(y_n | x_n)}{p(y_n | y_{1:n-1})}$$

A Motivating Example: Data



Example: Almost Constant Velocity Model

- States: $x_n = [s_n^x \; u_n^x \; s_n^y \; u_n^y]^T$
- Dynamics: $x_n = Ax_{n-1} + \epsilon_n$

$$\begin{bmatrix} s_n^x \\ u_n^x \\ s_n^y \\ u_n^y \end{bmatrix} = \begin{bmatrix} 1 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta t \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_{n-1}^x \\ u_{n-1}^x \\ s_{n-1}^y \\ u_{n-1}^y \end{bmatrix} + \epsilon_n$$

- Observation: $y_n = Bx_n + \nu_n$

$$\begin{bmatrix} r_n^x \\ r_n^y \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} s_n^x \\ u_n^x \\ s_n^y \\ u_n^y \end{bmatrix} + \nu_n$$

Sampling Approaches

The Monte Carlo Method

- Given a probability density, f, and $\varphi : E \to \mathbb{R}$

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

- Simple Monte Carlo solution:
  - Sample $X_1, \ldots, X_N \overset{\text{i.i.d.}}{\sim} f$.
  - Estimate $\widehat{I} = \frac{1}{N} \sum_{i=1}^{N} \varphi(X_i)$.
- Can also be viewed as approximating $\pi(dx) = f(x)dx$ with

$$\widehat{\pi}^N(dx) = \frac{1}{N} \sum_{i=1}^{N} \delta_{X_i}(dx).$$

MASDOC

The Importance–Sampling Identity

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

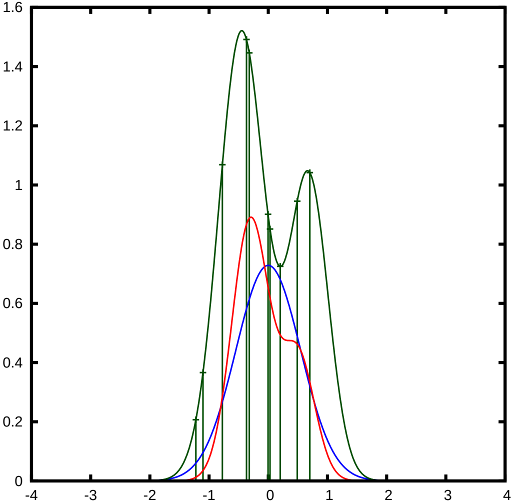
  define $w(x) = f(x)/g(x)$ and:

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

- This suggests the importance sampling estimator:
  - Sample $X_1, \ldots, X_N \overset{\text{i.i.d.}}{\sim} g$.
  - Estimate $\widehat{I} = \frac{1}{N} \sum_{i=1}^{N} w(X_i) \varphi(X_i)$.
- Can also be viewed as approximating $\pi(dx) = f(x)dx$ with

$$\widehat{\pi}^N(dx) = \frac{1}{N} \sum_{i=1}^{N} w(X_i) \delta_{X_i}(dx).$$

Importance Sampling Example



Self-Normalised Importance Sampling

- Often, f is known only up to a normalising constant.
- If $v(x) = cf(x)/g(x) = cw(x)$, then

$$\frac{\mathbb{E}_g(v\varphi)}{\mathbb{E}_g(v\mathbf{1})} = \frac{\mathbb{E}_g(cw\varphi)}{\mathbb{E}_g(cw\mathbf{1})} = \frac{c\mathbb{E}_f(\varphi)}{c\mathbb{E}_f(\mathbf{1})} = \mathbb{E}_f(\varphi).$$

- Estimate the numerator and denominator with the same sample:

$$\widehat{I} = \frac{\sum_{i=1}^{N} v(X_i)\varphi(X_i)}{\sum_{i=1}^{N} v(X_i)}.$$

- Biased for finite samples, but consistent.
- Typically reduces variance.

MASDOC

Importance Sampling for Smoothing/Filtering

- Sample $\{X_{1:n}^{(i)}\}$ at time n from $q_n(x_{1:n})$, define

$$w_n(x_{1:n}) \propto \frac{p(x_{1:n}|y_{1:n})}{q(x_{1:n})} = \frac{p(x_{1:n}, y_{1:n})}{q(x_{1:n})p(y_{1:n})}$$
$$\propto \frac{f(x_1)g(y_1|x_1)\prod_{m=2}^{n} f(x_m|x_{m-1})g(y_m|x_m)}{q_n(x_{1:n})}$$

- set $W_n^{(i)} = w_n(X_{1:n}^{(i)})/\sum_j w_n(X_{1:n}^{(j)})$,
- then $\{W_n^{(i)}, X_n^{(i)}\}$ is a consistently weighted sample.
- This seems inefficient.

Sequential Importance Sampling (SIS) I

- Importance weight

$$w_n(x_{1:n}) \propto \frac{f(x_1)g(y_1|x_1)\prod_{m=2}^{n} f(x_m|x_{m-1})g(y_m|x_m)}{q_n(x_{1:n})}$$

$$= \frac{f(x_1)g(y_1|x_1)}{q_n(x_1)} \prod_{m=2}^{n} \frac{f(x_m|x_{m-1})g(y_m|x_m)}{q_n(x_m|x_{1:m-1})}$$

- Given $\{W_{n-1}^{(i)}, X_{1:n-1}^{(i)}\}$ targetting $p(x_{1:n-1}|y_{1:n-1})$
  - Let $q_n(x_{1:n-1}) = q_{n-1}(x_{1:n-1})$,
  - sample $X_n^{(i)} \overset{\text{i.i.d.}}{\sim} q_n(\cdot|X_{n-1}^{(i)})$.

Sequential Importance Sampling (SIS) II

- And update the weights:

$$
\begin{aligned}
w_n(x_{1:n}) =& w_{n-1}(x_{1:n-1}) \frac{f(x_n|x_{n-1})g(y_n|x_n)}{q_n(x_n|x_{n-1})} \\
W_n^{(i)} =& w_n(X_{1:n}^{(i)}) \\
=& w_{n-1}(X_{1:n-1}^{(i)}) \frac{f(X_n^{(i)}|X_{n-1}^{(i)})g(y_n|X_n^{(i)})}{q_n(X_n^{(i)}|X_{n-1}^{(i)})} \\
=& W_{n-1}^{(i)} \frac{f(X_n^{(i)}|X_{n-1}^{(i)})g(y_n|X_n^{(i)})}{q_n(X_n^{(i)}|X_{n-1}^{(i)})}
\end{aligned}
$$

- If $\int p(x_{1:n}|y_{1:n})dx_n \approx p(x_{1:n-1}|y_{1:n-1})$ this makes sense.
- We only need to store $\{W_n^{(i)}, X_{n-1:n}^{(i)}\}$.
- Same computation every iteration.

MASDOC

Importance Sampling on Huge Spaces Doesn't Work

- It's said that IS *breaks the curse of dimensionality*:

$$\sqrt{N}\left[\frac{1}{N}\sum_{i=1}^{N} w(X_i)\varphi(X_i) - \int \varphi(x)f(x)dx\right] \stackrel{\mathrm{d}}{\to} \mathcal{N}(0, \mathbb{V}\mathrm{ar}_g\left[w\varphi\right])$$

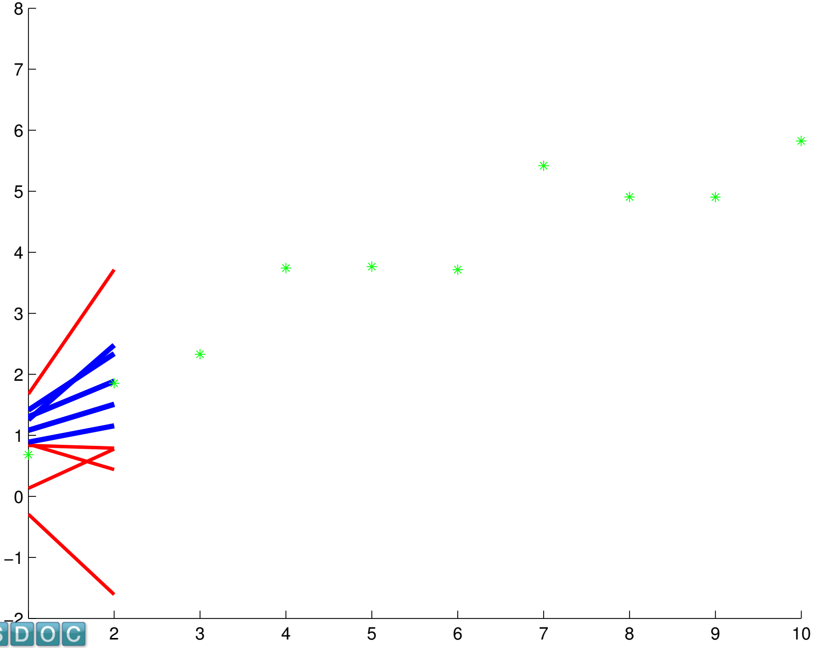
- This is true.
- But it's not *enough*.
- $\mathbb{V}\mathrm{ar}_g\left[w\varphi\right]$ increases (often exponentially) with dimension.
- **Eventually**, an SIS estimator (of $p(x_{1:n}|y_{1:n})$) **will** fail.
- But $p(x_n|y_{1:n})$ is a *fixed-dimensional* distribution.

Sequential Importance Resampling

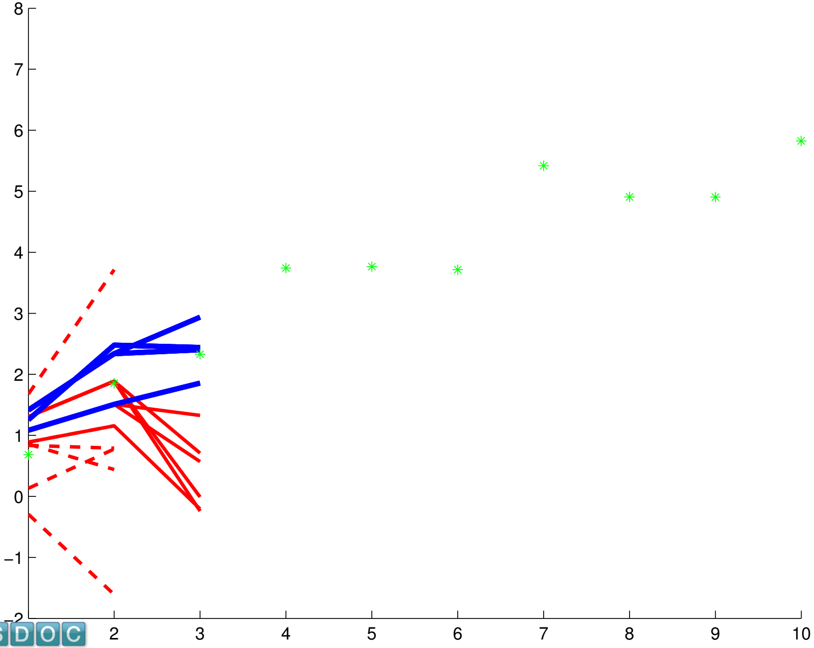
Resampling: The SIR[esampling] Algorithm

- Problem: variance of the weights builds up over time.
- Solution? Given $\{W_{n-1}^{(i)}, X_{1:n-1}^{(i)}\}$:
  1. Resample, to obtain $\{\frac{1}{N}, \widetilde{X}_{1:n-1}^{(i)}\}$.
  2. Sample $X_n^{(i)} \sim q_n(\cdot|\widetilde{X}_{n-1}^{(i)})$.
  3. Set $X_{1:n-1}^{(i)} = \widetilde{X}_{1:n-1}^{(i)}$.
  4. Set $W_n^{(i)} = f(X_n^{(i)}|X_{n-1}^{(i)})g(y_n|X_n^{(i)})/q_n(X_n^{(i)}|X_{n-1}^{(i)})$.
- And continue as with SIS.
- There is a cost, but this really works.

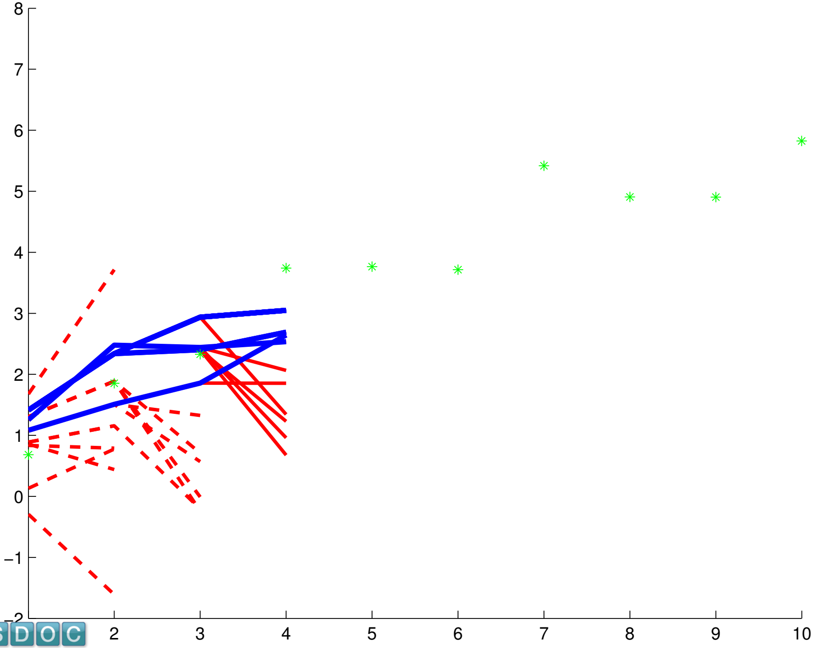
Iteration 2



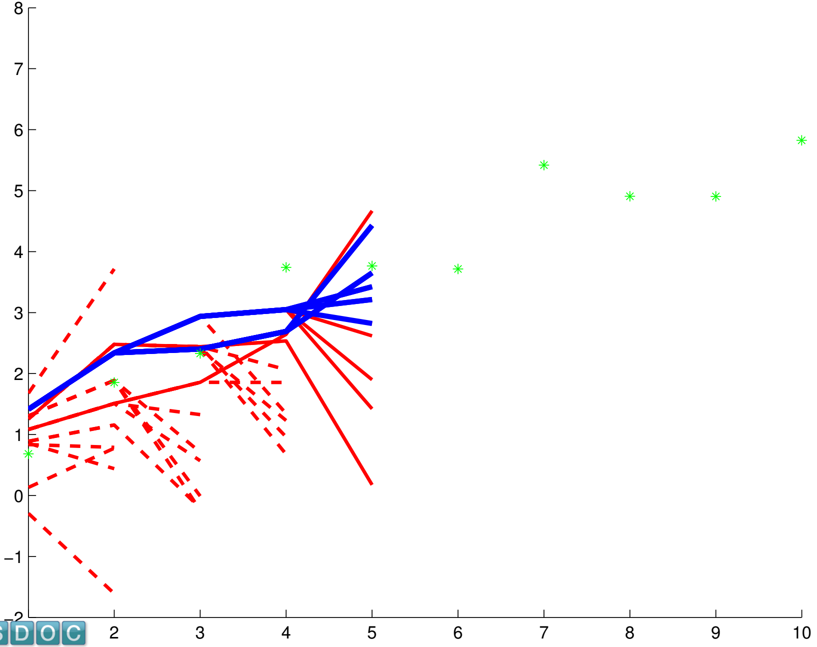
Iteration 3



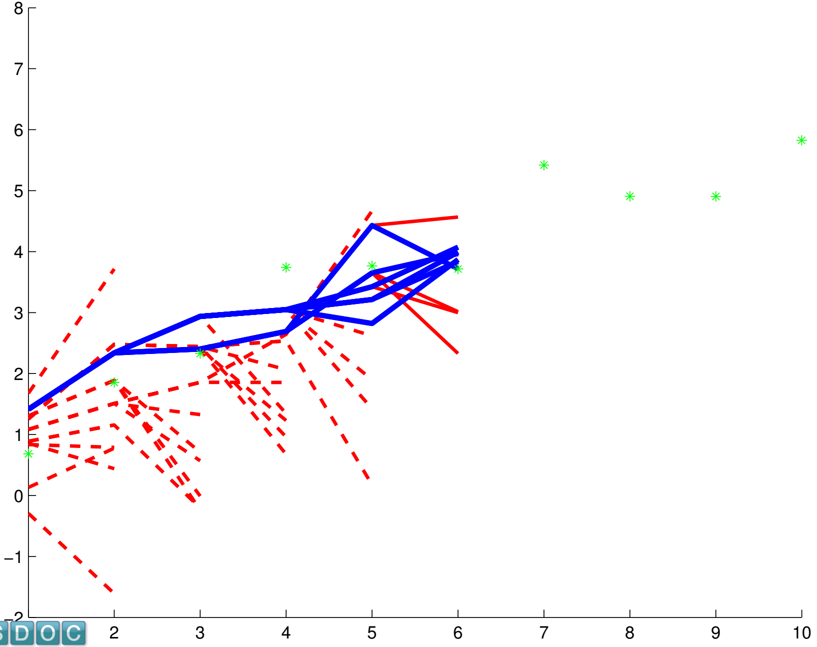
Iteration 4



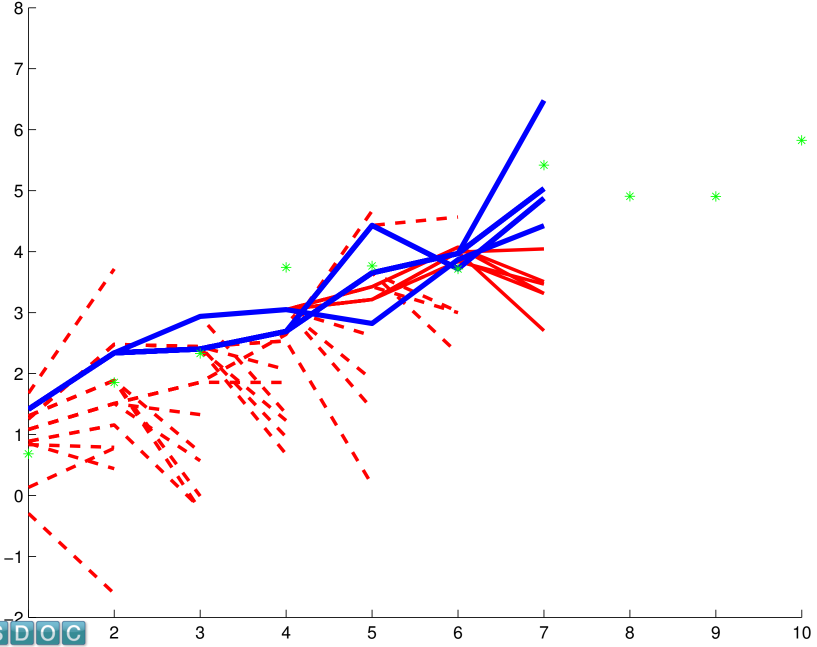
Iteration 5



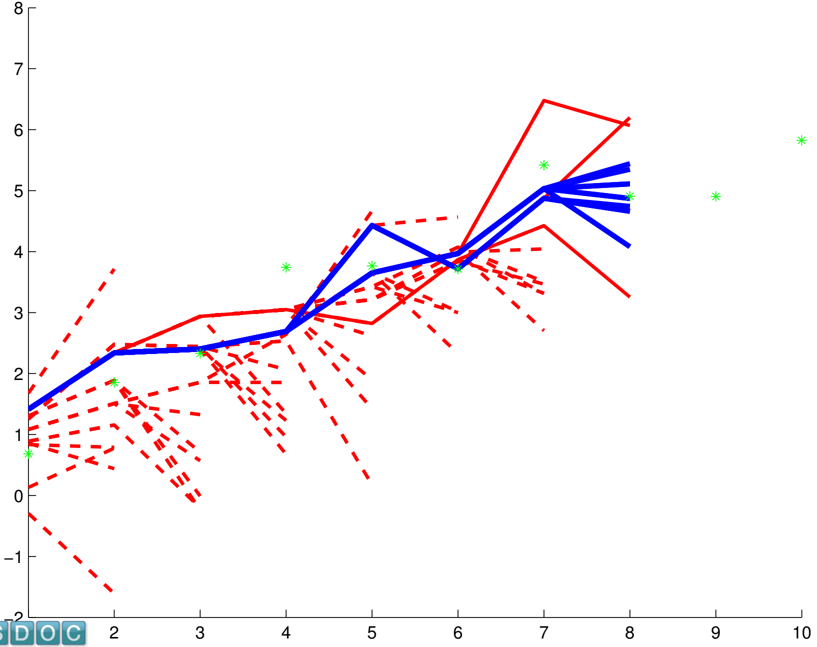
Iteration 6



Iteration 7

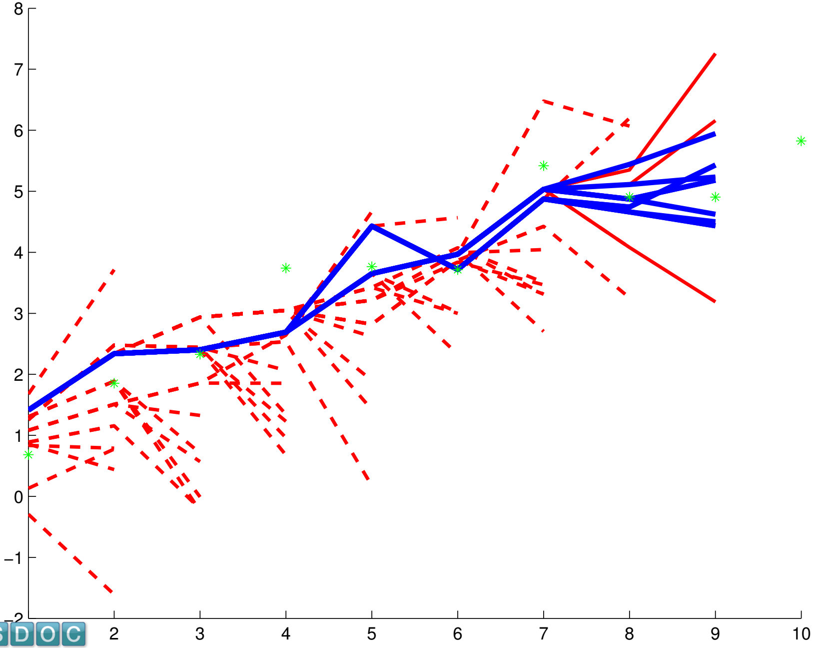


Iteration 8



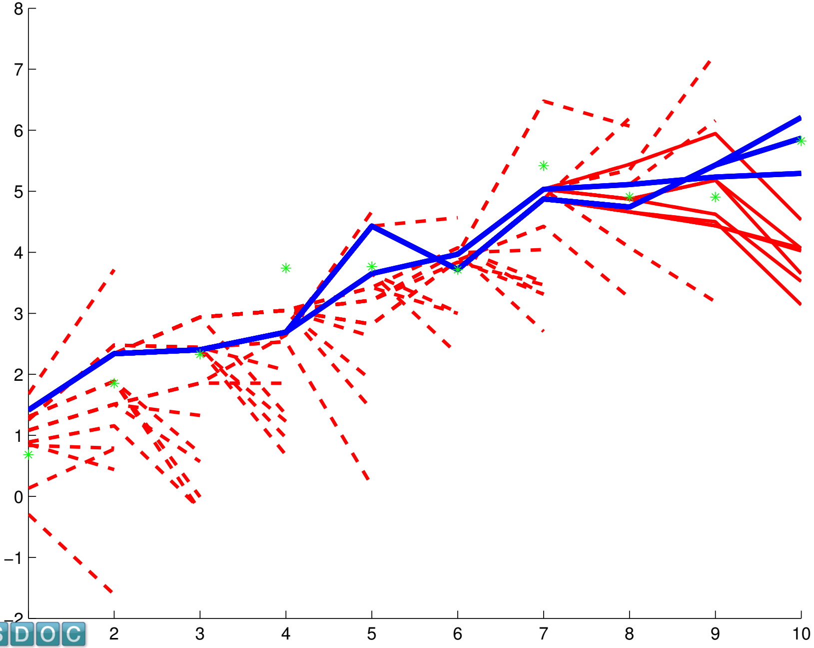
MASDOC

Iteration 9



MASDOC

Iteration 10



Part 2 – Applied Probability

Feynman-Kac Formulæ

Feynman-Kac Formulæ

- A natural description for measure-valued stochastic processes.
- Model for:
  - Particle motion in absorbing environments.
  - Classes of branching particle system.
  - Simple genetic algorithms.
  - Particle filters and related algorithms.

Structure of this section:

- Probabilistic Construction
- Semigroup[oid] Structure
- McKean Interpretations
- Particle Approximations
- Selected Results

Probabilistic Construction

The Canonical Markov Chain

- Consider the filtered probability space:

$$(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n\in\mathbb{N}}, \mathbb{P}_\mu)$$

- Let $\{X_n\}_{n\in\mathbb{N}}$ be a Markov chain such that for any $n \in \mathbb{N}$:

$$\mathbb{P}_\mu(X_{1:n} \in dx_{1:n}) = \mu(dx_1) \prod_{i=2}^{n} M_i(x_{i-1}, dx_i)$$

$$X_i : \Omega \to E_i \qquad \mu \in \mathcal{P}(E_1) \qquad M_i : E_{i-1} \to \mathcal{P}(E_i)$$

- $(E_i, \mathcal{E}_i)$ are measurable spaces.
- The X_i are $\mathcal{E}_i/\mathcal{F}_i$-measurable.
- Using Kolmogorov's/Tulcea's extension theorem there exists a unique process-valued extension.

Some Operator Notation

Given two measurable spaces, $(E, \mathcal{E})$ and $(F, \mathcal{F})$, a measure μ on $(E, \mathcal{E})$ and a Markov kernel, $K : E \to \mathcal{P}(F)$, define:

$$\mu(\varphi_E) := \int \mu(dx)\varphi_E(x)$$

$$\mu K(\varphi_F) := \int \mu(dx)K(x, dy)\varphi_F(y) \qquad \mu K \in \mathcal{P}(F)$$

$$K(\varphi_F)(x) := \int K(x, dy)\varphi_F(y) \qquad K(\varphi_F) : E \to \mathbb{R}$$

with φ_E, φ_F suitably measurable functions.
Given two functions, $f, g : E \to \mathbb{R}$, define $f \cdot g : E \to \mathbb{R}$ via $(f \cdot g)(x) = f(x)g(x)$.

The Feynman-Kac Formulæ

- Given $\mathbb{P}_\mu$ and **potential functions**:

$$\{G_i\}_{i\in\mathbb{N}} \qquad\qquad G_i : E_i \to [0,\infty)$$

- Define two path measures weakly:

$$\mathbb{Q}_n(\varphi_{1:n}) = \frac{\mathbb{E}\left[\varphi_{1:n}(X_{1:n})\prod_{i=1}^{n-1} G_i(X_i)\right]}{\mathbb{E}\left[\prod_{i=1}^{n-1} G_i(X_i)\right]}$$

$$\widehat{\mathbb{Q}}_n(\varphi_{1:n}) = \frac{\mathbb{E}\left[\varphi_{1:n}(X_{1:n})\prod_{i=1}^{n} G_i(X_i)\right]}{\mathbb{E}\left[\prod_{i=1}^{n} G_i(X_i)\right]}$$

where $\varphi_{1:n} : \otimes_{i=1}^{n} E_i \to \mathbb{R}$.

Example (Filtering via FK Formulæ: Prediction)

- Let $\mu(x_1) = f(x_1)$, $M_n(x_{n-1}, dx_n) = f(x_n|x_{n-1})dx_n$.
- Let $G_n(x_n) = g(y_n|x_n)$.
- Then:

$$\begin{aligned}
\mathbb{Q}_n(\varphi_{1:n}) =& \mathbb{E}\left[\varphi_{1:n}(X_{1:n}) \prod_{i=1}^{n-1} G_i(X_i)\right] \Big/ \mathbb{E}\left[\prod_{i=1}^{n-1} G_i(X_i)\right] \\
=& \mathbb{E}\left[\varphi_{1:n}(X_{1:n}) \prod_{i=1}^{n-1} g(y_i|X_i)\right] \Big/ \mathbb{E}\left[\prod_{i=1}^{n-1} g(y_i|X_i)\right] \\
=& \frac{\int \left[f(x_1) \prod_{i=2}^{n} f(x_i|x_{i-1})\right] \left[\prod_{j=1}^{n-1} g(y_j|x_j)\right] \varphi_{1:n}(x_{1:n}) dx_{1:n}}{\int \left[f(x_1) \prod_{i=2}^{n} f(x_i|x_{i-1})\right] \left[\prod_{j=1}^{n-1} g(y_j|x_j)\right] dx_{1:n}} \\
=& \int p(x_{1:n}|y_{1:n-1}) \varphi_{1:n}(x_{1:n}) dx_{1:n}
\end{aligned}$$

Example (Filtering via FK Formulæ: Update/Filtering)

- Whilst:

$$
\begin{aligned}
\widehat{\mathbb{Q}}_n\left(\varphi_{1:n}\right) =& \mathbb{E}\left[\varphi_{1:n}(X_{1:n})\prod_{i=1}^{n} G_i(X_i)\right] \Big/ \mathbb{E}\left[\prod_{i=1}^{n} G_i(X_i)\right] \\
=& \mathbb{E}\left[\varphi_{1:n}(X_{1:n})\prod_{i=1}^{n} g(y_i|X_i)\right] \Big/ \mathbb{E}\left[\prod_{i=1}^{n} g(y_i|X_i)\right] \\
=& \frac{\int \left[f(x_1)\prod_{i=2}^{n} f(x_i|x_{i-1})\right]\left[\prod_{j=1}^{n} g(y_j|x_j)\right]\varphi_{1:n}(x_{1:n})dx_{1:n}}{\int \left[f(x_1)\prod_{i=2}^{n} f(x_i|x_{i-1})\right]\left[\prod_{j=1}^{n} g(y_j|x_j)\right]dx_{1:n}} \\
=& \int p(x_{1:n}|y_{1:n})\varphi_{1:n}(x_{1:n})dx_{1:n}
\end{aligned}
$$

Feynman-Kac Marginal Measures

We are typically interested in marginals:

$$\gamma_n(\varphi_n) = \mathbb{E}\left[\varphi_n(X_n)\prod_{i=1}^{n-1} G_i(X_i)\right] \qquad \widehat{\gamma}_n(\varphi_n) = \mathbb{E}\left[\varphi_n(X_n)\prod_{i=1}^{n} G_i(X_i)\right]$$

$$\eta_n(\varphi_n) = \mathbb{Q}_n(\mathbf{1}_{1:n-1} \otimes \varphi_n) \qquad \widehat{\eta}_n = \widehat{\mathbb{Q}}_n(\mathbf{1}_{1:n-1} \otimes \varphi_n)$$

$$= \frac{\mathbb{E}\left[\varphi_n(X_n)\prod_{i=1}^{n-1} G_i(X_i)\right]}{\mathbb{E}\left[\prod_{i=1}^{n-1} G_i(X_i)\right]} \qquad = \frac{\mathbb{E}\left[\varphi_n(X_n)\prod_{i=1}^{n} G_i(X_i)\right]}{\mathbb{E}\left[\prod_{i=1}^{n} G_i(X_i)\right]}$$

$$= \gamma_n(\varphi_n)/\gamma_n(\mathbf{1}) \qquad = \widehat{\gamma}_n(\varphi_n)/\widehat{\gamma}_n(\mathbf{1})$$

Key property:

$$\eta_n(A_n) = \int_{E_1 \times \ldots E_{n-1} \times A_n} \mathbb{Q}_n(dx_{1:n})$$

$$\widehat{\eta}_n(A_n) = \int_{E_1 \times \ldots E_{n-1} \times A_n} \widehat{\mathbb{Q}}_n(dx_{1:n})$$

Analysis: Semigroup Structure

How do the marginal distributions evolve?

Some Recursive Relationships

- The unnormalized marginals obey:

$$\widehat{\gamma}_n(\varphi_n) = \gamma_n(\varphi_n \cdot G_n) \qquad \gamma_n(\varphi_n) = \widehat{\gamma}_{n-1} M_n(\varphi_n)$$

- Whilst the normalized marginals satisfy:

$$\begin{aligned} \widehat{\eta}_n(\varphi_n) &= \frac{\widehat{\gamma}_n(\varphi_n)}{\widehat{\gamma}_n(\mathbf{1})} \\ &= \frac{\gamma_n(\varphi_n \cdot G_n)}{\gamma_n(G_n)} \\ &= \frac{\eta_n(\varphi_n \cdot G_n)}{\eta_n(G_n)} \end{aligned} \qquad \begin{aligned} \eta_n(\varphi_n) &= \frac{\gamma_n(\varphi_n)}{\gamma_n(\mathbf{1})} \\ &= \frac{\widehat{\gamma}_{n-1} M_n(\varphi_n)}{\widehat{\gamma}_{n-1} M_n(\mathbf{1})} \\ &= \frac{\widehat{\eta}_{n-1} M_n(\varphi_n)}{\widehat{\eta}_{n-1} M_n(\mathbf{1})} \\ &= \widehat{\eta}_{n-1} M_n(\varphi_n) \end{aligned}$$

MASDOC

The Boltzmann-Gibbs Operator

- Given $\nu \in \mathcal{P}(E)$ and $G : E \to \mathbb{R}$:

$$\begin{aligned} \Psi_G : \quad & \mathcal{P}(E) \to \mathcal{P}(E) \\ \Psi_G : \quad & \nu \to \Psi_G(\nu) \end{aligned}$$

- The **Boltzmann-Gibbs** Operator Ψ_G is defined weakly by:

$$\forall \varphi \in \mathcal{C}_b : \qquad \Psi_G(\nu)(\varphi) = \frac{\nu(G \cdot \varphi)}{\nu(G)}$$

- or equivalently, for all measurable sets A:

$$\begin{aligned} \Psi_G(A) &= \frac{\nu(G \cdot \mathbb{I}_A)}{\nu(G)} \\ &= \frac{\int_A \nu(dx) G(x)}{\int_E \nu(dx') G(x')} \end{aligned}$$

Example (Boltzmann-Gibbs Operators and Bayes' Rule)

- Let $\mu(dx) = f(x)\lambda(dx)$ be a prior measure.
- Let $G(x) = g(y|x)$ be the likelihood.
- Then:

$$\begin{aligned}\Psi_G(\mu)(\varphi) &= \frac{\mu(G\cdot\varphi)}{\mu(G)} = \frac{\int \mu(dx)G(x)\varphi(x)}{\int \mu(dx')G(x')} \\ &= \frac{\int f(x)g(y|x)\varphi(x)\lambda(dx)}{\int f(x')g(y|x')\lambda(dx')} \\ &= \int f(x|y)\varphi(x)\lambda(dx)\end{aligned}$$

with

$$f(x|y) := \frac{f(x)g(y|x)}{\int f(x)g(y|x)\lambda(dx)}$$

- So: $\Psi_{g(y|\cdot)}$: Prior $\rightarrow$ Posterior.

Markov Semigroups

- A **semigroup** $\mathcal{S}$ comprises:
  - A set, S.
  - An associative binary operation, $\cdot$.
- A Markov Chain with homogeneous transition M has **dynamic semigroup** M_n:
  - $M_0(x, A) = \delta_x(A)$.
  - $M_1(x, A) = M(x, A)$.
  - $M_n(x, A) = \int M(x, dy) M_{n-1}(y, A)$.
  - $(M_n \cdot M_m)(x, A) = \int M_n(x, dy) M_m(y, A) = M_{n+m}(x, A)$.
- A *linear* semigroup.
- Key property:

$$\mathbb{P}\left(X_{n+m} \in A | X_m = x\right) = M_n(x, A).$$

Markov Semigroupoids

- A **semigroupoid**, $\mathcal{S}'$ comprises:
  - A set, S.
  - A *partial* associative binary operation, $\cdot$.
- A Markov Chain with inhomogeneous transitions M_n has **dynamic semigroupoid** $M_{p:q}$:
  - $M_{p:p}(x, A) = \delta_x(A)$.
  - $M_{p:p+1}(x, A) = M_{p+1}(x, A)$.
  - $M_{p:q}(x, A) = \int M_{p+1}(x, dy) M_{p+1:q}(y, A)$.
  - $(M_{p:q} \cdot M_{q:r})(x, A) = \int M_{p:q}(x, dy) M_{q:r}(y, A) = M_{p:r}(x, A)$.
- A *linear* semigroupoid.
- Key property:

$$\mathbb{P}\left(X_{n+m} \in A | X_m = x\right) = M_{m,n+m}(x, A).$$

MASDOC

An Unnormalized Feynman-Kac Semigroupoid

- We previously established:

$$\gamma_n = \widehat{\gamma}_{n-1} M_n \qquad \widehat{\gamma}_n(\varphi_n) = \gamma_n(\varphi_n \cdot G_n)$$

- Defining

$$Q_p(x_{p-1}, dx_p) = M_p(x_{p-1}, dx_p) G_p(x_p)$$

  we obtain $\gamma_n = \gamma_{n-1} Q_n$.
- We can construct the dynamic semigroupoid $Q_{p:q}$:
  - $Q_{p:p}(x, A) = \delta_x(A)$.
  - $Q_{p:p+1}(x, A) = Q_{p+1}(x, A)$.
  - $Q_{p:q}(x, A) = \int Q_{p+1}(x, dy) Q_{p+1:q}(y, A)$.
  - $(Q_{p:q} \cdot Q_{q:r})(x, A) = \int Q_{p:q}(x, dy) Q_{q:r}(y, A) = Q_{p:r}(x, A)$.
- *Just a Markov semigroupoid for general measures:*
  $\forall p \leq q : \gamma_q = \gamma_p Q_{p:q}$.

MASDOC

A Normalised Feynman-Kac Semigroupoid

- We previously established:

$$\eta_n = \widehat{\eta}_{n-1} M_n(\varphi_n) \qquad \widehat{\eta}_n = \frac{\eta_n(\varphi_n \cdot G_n)}{\eta_n(G_n)}$$

- From the definition of Ψ_{G_n}: $\widehat{\eta}_n = \Psi_{G_n}(\eta_n)$.
- Defining $\Phi_n : \mathcal{P}(E_{n-1}) \to \mathcal{P}(E_n)$ as:

$$\Phi_n : \eta_{n-1} \to \Psi_{G_{n-1}}(\eta_{n-1}) M_n$$

  we have the recursion $\eta_n = \Phi_n(\eta_{n-1})$ and the nonlinear semigroupoid, $\Phi_{p:q}$:
    - $\Phi_{p:p}(x, A) = \delta_x(A)$.
    - $\Phi_{p:p+1}(x, A) = \Phi_{p+1}(x, A)$.
    - $\Phi_{p:q}(x, A) = \Phi_{p+1:q}(\Phi_{p+1}(\eta_p))$ for $q > p + 1$.
    - $(\Phi_{p:q} \cdot \Phi_{q:r})(x, A) = \int \Phi_{q:r}(y, A) \Phi_{p:q}(x, dy) = \Phi_{p:r}(x, A)$.
- Again: $\forall p \leq q : \ \eta_q = \eta_p \Phi_{p:q}$.

MASDOC

McKean Interpretations

Microscopic mass transport.

McKean Interpretations of Feynman-Kac Formulæ

- Families of Markov kernels consistent with FK Marginals.
- A collection $\{K_{n,\eta}\}_{n\in\mathbb{N},\eta\in\mathcal{P}(E_{n-1})}$ is a **McKean Interpretation** if:

$$\forall n \in \mathbb{N}: \ \eta_n = \Phi_n(\eta_{n-1}) = \eta_{n-1}K_{n,\eta_{n-1}}.$$

- Not unique…and not linear.
- Selection/Mutation approach seems natural:
  - Choose $S_{n,\eta}$ such that $\eta S_{n,\eta} = \Psi_{G_n}(\eta)$.
  - Set $K_{n+1,\eta} = S_{n,\eta}M_{n+1}$.
- Still not unique:
  - $S_{n,\eta}(x_n,\cdot) = \Psi_{G_n}(\eta)$
  - $S_{n,\eta}(x_n,\cdot) = \epsilon_n G_n(x_n)\delta_{x_n}(\cdot) + (1-\epsilon_n G_n(x_n))\Psi_{G_n}(\eta)(\cdot)$

Particle Interpretations

Stochastic discretisations.

Particle Interpretations of Feynman-Kac Formulæ I

Given a McKean interpretation, we can attach an N-particle model.

- Denote $\xi_n^{(N)} = (\xi_n^{(N,1)}, \xi_n^{(N,2)}, \ldots, \xi_n^{(N,N)}) \in E_n^N$.
- Allow
$$\left(\Omega^N, \mathcal{F}^N = (\mathcal{F}_n^N)_{n\in\mathbb{N}}, \xi^{(N)}, \mathbb{P}_{\eta_0}^N\right)$$
to indicate a particle-set-valued Markov chain.
- Let $\eta_{n-1}^{(N)} = \frac{1}{N}\sum_{i=1}^N \delta_{\xi_{n-1}^{(N,i)}}$.
- Allow the elementary transitions to be:

$$\mathbb{P}\left(\xi_n^{(N)} \in d\xi_n^{(N)} | \xi_{n-1}^{(N)}\right) = \prod_{p=1}^{N} K_{n,\eta_{n-1}^{(N)}}(\xi_{n-1}^{(N,p)}, d\xi_n^{(N,p)})$$

Particle Interpretations of Feynman-Kac Formulæ II

- Consider $K_{n,\eta} = S_{n-1,\eta} M_n$

$$\mathbb{P}\left(\xi_n^{(N)} \in d\xi_n^{(N)} | \xi_{n-1}^{(N)}\right) = \prod_{p=1}^{N} S_{n-1,\eta_{n-1}^{(N)}} M_n(\xi_{n-1}^{(N,p)}, d\xi_n^{(N,p)})$$

- Defining:

$$\mathcal{S}_{n-1}^{(N)}(\xi_{n-1}^{(N)}, d\widehat{\xi}_n^{(N)}) = \prod_{i=1}^{N} S_{n,\eta_{n-1}^{(N)}}(\xi_{n-1}^{(N,p)}, d\widehat{\xi}_{n-1}^{(N,p)})$$

$$\mathcal{M}_n^{(N)}(\widehat{\xi}_{n-1}^{(N)}, d\xi_n^{(N)}) = \prod_{i=1}^{N} M_n(\widehat{\xi}_{n-1}^{(N,p)}, \xi_n^{(N,p)})$$

it is clear that:

$$\mathbb{P}\left(\xi_n^{(N)} \in d\xi_n^{(N)} | \xi_{n-1}^{(N)}\right) = \int_{E_{n-1}^N} \mathcal{S}_{n-1,\eta_{n-1}^{(N)}}(\xi_{n-1}^{(N)}, d\widehat{\xi}_{n-1}^{(N)}) \mathcal{M}_n(\widehat{\xi}_{n-1}^{(N)}, d\xi_n^{(N)})$$

Selection, Mutation and Structure

- A suggestive structural similarity:

$$\eta_{n-1} \in \mathcal{P}(E_{n-1}) \xrightarrow{S_{n-1,\eta_{n-1}}} \widehat{\eta}_n \in \mathcal{P}(E_{n-1}) \xrightarrow{M_n} \eta_n \in \mathcal{P}(E_n)$$

$$\xi_{n-1}^{(N)} \in E_{n-1}^N \xrightarrow{\text{Select}} \widehat{\xi}_n^{(N)} \in E_{n-1}^N \xrightarrow{\text{Mutate}} \xi_n^{(N)} \in E_n^N$$

- Selection:

$$S_{n-1,\eta_{n-1}^{(N)}} = \Psi_{G_{n-1}}(\eta_{n-1}^{(N)}) = \sum_{i=1}^{N} \frac{G_{n-1}(\xi_{n-1}^{(N,i)})}{\sum_{j=1}^{N} G_{n-1}(\xi_{n-1}^{(N,j)})} \delta_{\xi_{n-1}^{(N,i)}}$$

$$\widehat{\xi}_{n-1}^{(N,i)} \overset{\text{i.i.d.}}{\sim} \Psi_{G_{n-1}}(\eta_{n-1}^{(N)})$$

- Mutation:

$$\xi_n^{(N,i)} \overset{\text{i.i.d.}}{\sim} M_n(\widehat{\xi}_{n-1}^{(N,i)}, d\xi_n^{(N,i)})$$

- Semigroupoid

$$\mathbb{P}^N(\xi_n^{(N)} \in dx_n^{(N)} | \xi_{n-1}^{(N)}) = \prod_{i=1}^{N} \Phi_n(\eta_{n-1}^{(N)})(dx_n^{(N,i)})$$

Selected Results

Local Error Decomposition

$$
\begin{array}{ccccccccc}
\eta_1 & \to & \eta_2 = \Phi_2(\eta_1) & \to & \eta_3 = \Phi_{1:3}(\eta_1) & \to & \dots & \to & \Phi_{1:n}(\eta_1) \\
\Downarrow & & & & & & & & \\
\eta_1^N & \to & \Phi_2(\eta_1^N) & \to & \Phi_{1:3}(\eta_1^N) & \to & \dots & \to & \Phi_{1:n}(\eta_1^N) \\
 & & \Downarrow & & & & & & \\
 & & \eta_2^N & \to & \Phi_3(\eta_2^N) & \to & \dots & \to & \Phi_{2:n}(\eta_2^N) \\
 & & & & \Downarrow & & & & \\
 & & & & \eta_3^N & \to & \dots & \to & \Phi_{3:n}(\eta_3^N) \\
 & & & & & & \vdots & & \\
 & & & & & & \Downarrow & & \\
 & & & & & & \eta_{n-1}^N & \to & \Phi_n(\eta_{n-1}^N) \\
 & & & & & & & & \Downarrow \\
 & & & & & & & & \eta_n^N
\end{array}
$$

Formally: $\eta_n^N - \eta_n = \sum_{i=1}^{n} \Phi_{i,n}(\eta_i^N) - \Phi_{i,n}(\Phi_i(\eta_{i-1}^N))$

MASDOC

A Key Martingale

Proposition (Del Moral, 2004 Propostion 7.4.1)

For each $n \geq 0$, $\varphi_n \in \mathcal{C}_b(E_n)$ *define:*

$$\Gamma^N_{\cdot,n}(\varphi_n) : p \in \{1, \ldots, n\} \to \Gamma_{p,n}(\varphi_n)$$
$$\Gamma^N_{p,n}(\varphi_n) := \gamma^N_p(Q_{p,n}\varphi_n) - \gamma_p(Q_{p,n}\varphi_n)$$

For any $p \leq n$: $\Gamma^N_{\cdot,n}(\varphi_n)$ *has* $\mathcal{F}^N$*-martingale decomposition:*

$$\Gamma^N_{p,n}(\varphi_n) = \sum_{q=1}^{p} \gamma^N_q(\mathbf{1}) \left[\eta^N_q(Q_{q,n}\varphi_n) - \eta^N_{q-1} K_{q,\eta^N_{q-1}}(Q_{q,n}\varphi_n) \right]$$

$$\left\langle \Gamma^N_{\cdot,n} \right\rangle_p = \sum_{q=1}^{p} \gamma^N_q(\mathbf{1})^2 \eta^N_{q-1} \left[K_{q,\eta^N_{q-1}}(Q_{q,n}(\varphi_n) - \eta^N_{q-1} K_{q,\eta^N_{q-1}} Q_{q,n}(\varphi_n)) \right]^2$$

MASDOC

Normalizing the Unnormalized

$$
\begin{aligned}
\eta_n^N(\varphi_n) - \eta_n(\varphi_n) &= \frac{\gamma_n^N(\varphi_n)}{\gamma_n^N(\mathbf{1})} - \frac{\gamma_n(\varphi_n)}{\gamma_n(\mathbf{1})} \\
&= \frac{\gamma_n(\mathbf{1})}{\gamma_n^N(\mathbf{1})} \left[\frac{\gamma_n^N(\varphi_n)}{\gamma_n(\mathbf{1})} - \frac{\gamma_n(\varphi_n)}{\gamma_n(\mathbf{1})} \times \frac{\gamma_n^N(\mathbf{1})}{\gamma_n(\mathbf{1})} \right] \\
&= \frac{\gamma_n(\mathbf{1})}{\gamma_n^N(\mathbf{1})} \left[\frac{\gamma_n^N(\varphi_n)}{\gamma_n(\mathbf{1})} - \eta_n(\varphi_n) \times \frac{\gamma_n^N(\mathbf{1})}{\gamma_n(\mathbf{1})} \right] \\
&= \frac{\gamma_n(\mathbf{1})}{\gamma_n^N(\mathbf{1})} \left[\gamma_n^N \left(\frac{\varphi_n - \eta_n(\varphi_n)}{\gamma_n(\mathbf{1})} \right) \right]
\end{aligned}
$$

Law of Large Numbers and Weak Convergence

Theorem (Del Moral 2004: Theorem 7.4.4)

Under regularity conditions, for any $n \geq 1$, $p \geq 1$, $\varphi_n \in \mathcal{C}_b(E_n)$:

$$\sqrt{N}\mathbb{E}\left[|\eta_n^N(\varphi_n) - \eta_n(\varphi_n)|^p\right]^{1/p} \leq c_{p,n}||\varphi_n||_\infty$$

By a Borel-Cantelli argument:

$$\lim_{N\to\infty} \eta_n^N(\varphi_n) \stackrel{a.s.}{\to} \eta_n(\varphi_n).$$

Central Limit Theorem

Proposition (Del Moral 2004: Proposition 9.4.2)

Under regularity conditions, for any $n \geq 1$:

$$\sqrt{N}(\eta_n^N(\varphi_n) - \eta_n(\varphi_n)) \xrightarrow{d} \mathcal{N}\left(0, \sigma_n^2(\varphi_n)\right)$$

where

$$\sigma_n^2(\varphi_n) = \sum_{q=1}^{n} \eta_{q-1}\left[K_{q,\eta_{q-1}}(Q_{q,n}(\varphi_n) - K_{q,\eta_{q-1}}(Q_{q,n}(\varphi_n)))\right]^2$$

Part 3 – Interface
Particle Filters as McKean Interpretations

The Bootstrap Particle Filter

The Simplest Case

Recall: The SIR Particle Filter

- At iteration n, given $\{W_{n-1}^{(i)}, X_{1:n-1}^{(i)}\}$:
  1. Resample, to obtain $\{\frac{1}{N}, \widetilde{X}_{1:n-1}^{(i)}\}$. **Selection**
  2. Sample $X_n^{(i)} \sim q_n(\cdot|\widetilde{X}_{n-1}^{(i)})$. **Mutation**
  3. Set $X_{1:n-1}^{(i)} = \widetilde{X}_{1:n-1}^{(i)}$.
  4. Set $W_n^{(i)} = f(X_n^{(i)}|X_{n-1}^{(i)})g(y_n|X_n^{(i)})/q_n(X_n^{(i)}|X_{n-1}^{(i)})$.
- Feynman-Kac formulation?
  - Generally $W_n^{(i)}$ depends upon $X_{n-1}^{(i)}$.
  - (At least) 2 solutions exist.

The Bootstrap SIR Filter (Gordon, Salmond and Smith, 1993)

- The bootstrap particle filter:
  - Proposal: $q(x_{t-1}, x_t) = f(x_t|x_{t-1})$
  - Weight: $w(x_t) \propto g(y_t|x_t)$
- Feynman-Kac model:
  - Mutation: $M_t(x_{t-1}, dx_t) = f(x_t|x_{t-1})dx_t$.
  - Potential: $G_t(x_t) = g(y_t|x_t)$.
- McKean interpretation:
  - McKean transitions: $K_{n+1,\eta} = S_{n,\eta}M_{n+1}$.
  - Selection operation: $S_{n,\eta} = \Psi_{G_n}(\eta)$.

Bootstrap Particle Filter Results

LLN

$$\lim_{N\to\infty} \frac{\sum_{i=1}^{N} W_n^i \varphi_n(X_n^i)}{\sum_{j=1}^{N} W_n^j} \xrightarrow{\text{a.s.}} \int \varphi_n(x_n) p(x_n|y_{1:n}) dx_n$$

CLT

$$\sqrt{N}\left(\frac{\sum_{i=1}^{N} W_n^i \varphi_n(X_n^i)}{\sum_{j=1}^{N} W_n^j} - \int \varphi_n(x_n) p(x_n|y_{1:n}) dx_n\right)$$

$$\xrightarrow{\text{d}} \mathcal{N}\left(0, \sigma_{BS,n}^2\left(\varphi_n\right)\right)$$

Bootstrap Particle Filter: Asymptotic Variance

$$
\begin{aligned}
&\sigma^2_{BS,n}(\varphi_n) = \\
&\int \frac{p(x_1|y_{1:n})^2}{p(x_1)} \left(\int \varphi_n(x_n) p(x_n|y_{2:n}, x_1) dx_n - \bar{\varphi}_n \right)^2 dx_1 \\
&+ \sum_{k=2}^{t-1} \int \frac{p(x_k|y_{1:n})^2}{\mathbf{p(x_k|y_{1:k-1})}} \left(\int \varphi_n(x_n) p(x_n|y_{k+1:n}, x_k) dx_n - \bar{\varphi}_n \right)^2 dx_{1:k} \\
&+ \int \frac{p(x_n|y_{1:n})^2}{\mathbf{p(x_n|y_{1:n-1})}} \left(\varphi_n(x_n) - \bar{\varphi}_n \right)^2 dx_n.
\end{aligned}
$$

with

$$
\bar{\varphi}_n = \int p(x_n|y_{1:n}) \varphi_n(x_n) dx_n.
$$

Extended Spaces: General SIR Particle Filter

- At iteration n, given $\{W_{n-1}^{(i)}, X_{1:n-1}^{(i)}\}$:
  1. Resample, to obtain $\{\frac{1}{N}, \widetilde{X}_{1:n-1}^{(i)}\}$. **Selection**
  2. Sample $X_n^{(i)} \sim q_n(\cdot|\widetilde{X}_{n-1}^{(i)})$. **Mutation**
  3. Set $X_{1:n-1}^{(i)} = \widetilde{X}_{1:n-1}^{(i)}$.
  4. Set $W_n^{(i)} = f(X_n^{(i)}|X_{n-1}^{(i)})g(y_n|X_n^{(i)})/q_n(X_n^{(i)}|X_{n-1}^{(i)})$.
- But $W_n^{(i)}$ depends upon $X_{n-1}^{(i)}$
- Let $\tilde{E}_n = E_{n-1} \times E_n$.
- Define $Y_n = (X_{n-1}, X_n)$.
- Now $W_n = \tilde{G}_n(Y_n)$.
- Set $\tilde{M}_n(y_{n-1}, dy_n) = \delta_{y_{n-1,2}}(dy_{n,1})q(y_{n-1,2}, dy_{n,1})$.
- A Feynman-Kac representation.

SIR Asymptotic Variance

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

Auxiliary Particle Filters

Another algorithm.

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 $\lambda_n^{(i)} \propto \widehat{p}(y_n|X_{n-1}^{(i)})$.
- Propose new states, from the mixture distribution

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

- Weight samples, correcting for the pre-weighting.

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

- Resample particle set.

MASDOC

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 $\lambda_n^{(i)} \propto \widehat{p}(y_n | X_{n-1}^{(i)})$.
- Resample
- Propose new states
- Weight samples, correcting for the pre-weighting.

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

A Feynman-Kac Interpretation of the APF

A transition and a potential.

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 $\lambda_{n+1}^{(i)} \propto \widehat{p}(y_{n+1}|X_n^{(i)})$.

An SIR algorithm targetting:

$$\widehat{p}_n(x_{1:n}|y_{1:n+1}) \propto p(x_{1:n}|y_{1:n})\widehat{p}(y_{n+1}|x_n).$$

Some Consequences

Asymptotics.

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

$$\sqrt{N}\left(\widehat{\varphi}^{N}_{n,APF} - \overline{\varphi}_n\right) \to \mathcal{N}\left(0, \sigma_n^2\left(\varphi_n\right)\right)$$

where,

$$\sigma_n^2(\varphi_n) =$$

$$\int \frac{p(x_1|y_{1:n})^2}{q_1(x_1)}\left(\int \varphi_n(x_n)p(x_n|y_{2:n},x_1)dx_n - \bar{\varphi}_n\right)^2 dx_1$$

$$+\sum_{k=2}^{t-1}\int \frac{p(x_{1:k}|y_{1:n})^2}{\color{red}{\widehat{\mathbf{p}}(\mathbf{x_{1:k-1}}|\mathbf{y_{1:k}})}q_k(x_k|x_{k-1})}\left(\int \varphi_n(x_n)p(x_n|y_{k+1:n},x_k)dx_n - \bar{\varphi}_n\right)^2 d$$

$$+\int \frac{p(x_{1:n}|y_{1:n})^2}{\color{red}{\widehat{\mathbf{p}}(\mathbf{x_{1:n-1}}|\mathbf{y_{1:n}})}q_n(x_n|x_{n-1})}\left(\varphi_n(x_n) - \bar{\varphi}_n\right)^2 dx_{1:n}.$$

Practical Implications

- It means we're doing importance sampling.
- Choosing $\widehat{p}(y_n|x_{n-1}) = p\left(y_n|x_n = \mathbb{E}\left[X_n|x_{n-1}\right]\right)$ 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

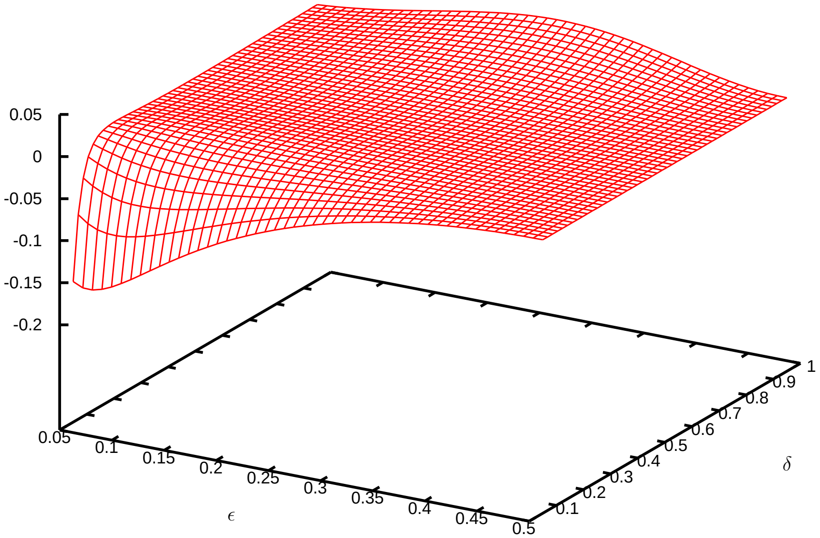
Consider the following binary state-space model with common state and observation spaces:

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

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

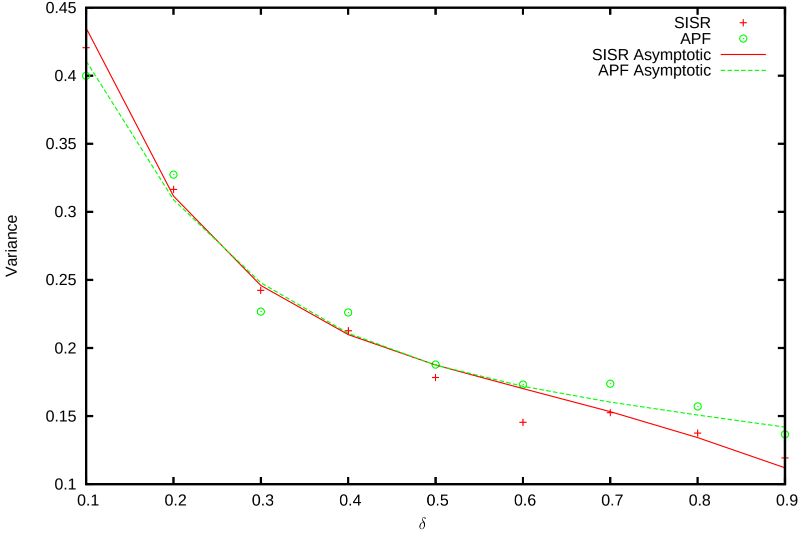
Consider estimating $\mathbb{E}(X_2 | Y_{1:2} = (0,1))$.

Variance of SIR - Variance of APF



MASDOC

Variance Comparison at $\epsilon = 0.25$



MASDOC

Further Reading

Particle Filters / Sequential Monte Carlo

- Novel approach to nonlinear/non-Gaussian Bayesian state estimation, Gordon, Salmond and Smith, *IEE Proceedings-F*, **140**(2):107-113, 1993.
- *Sequential Monte Carlo in Practice*, Doucet, De Freitas & Gordon, Springer, 2001.
- A survey of convergence results on particle filtering methods for practitioners., Crisan and Doucet, *IEEE Transactions on Signal Processing*, textbf50(3):736-746, 2002.
- Central limit theorem for sequential Monte Carlo methods and its applications to Bayesian inference. Chopin, *Annals of Statistics*, **32**(6):2385-2411, 2004.
- A Tutorial on Particle Filtering. . . , Doucet & **J.**, in *Oxford Handbook of Nonlinear Filtering*, 2010 (to appear).

- Sequential Monte Carlo Samplers, Doucet, Del Moral & Jasra, *JRSSB*, **63**(3):411-436, 2006.

Feynman-Kac Particle Models

- *Particle Methods: An Introduction with Applications*, Del Moral & Doucet, HAL INRIA Report RR-6691, 2009.
- *Feynman-Kac Formulæ...*, Del Moral, Springer, 2004.

Auxiliary Particle Filters

- Filtering Via Simulation, Pitt & Shepherd, *JASA*, **94**(446):590-599, 1999.
- A Note on Auxiliary Particle Filters, **J.** & Doucet, *Stat. & Prob. Lett.*, **78**(12):1498–1504, 2008.
- Optimality of the APF, Douc, Moulines & Olsson, *Prob. & Math. Stat.*, **29**(1):1–28, 2009.