Sequential Monte Carlo and Integral Equations

Adam M. Johansen

a.m.johansen@warwick.ac.uk
http://go.warwick.ac.uk/amjohansen/talks/

ST912 Statistical Frontiers Seminar
March 2021

Warwick Statistics

Overview

Motivating Problem Filtering

Monte Carlo Solutions Sequential Monte Carlo algorithms

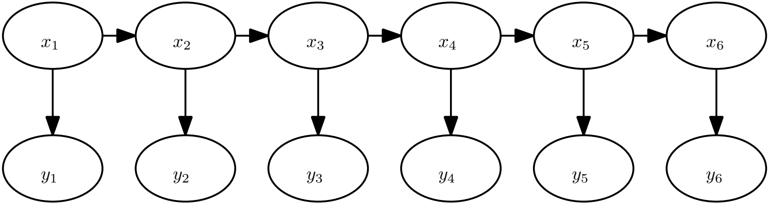
Feynman-Kac Formulae A mathematical description

Fredholm Equations A non-standard application

Filtering

A Motivating Problem

(General State Space) Hidden Markov Models / State Space Models



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

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

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:

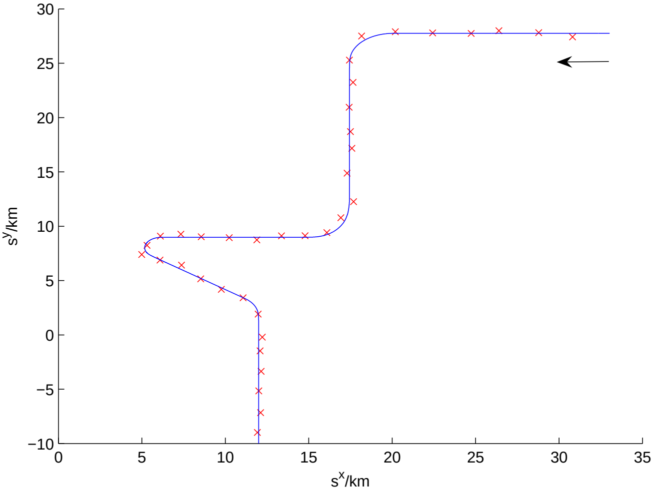
$$Y_n | \{X_n = x_n\} \sim g(\cdot | x_n).$$

- 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:

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

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

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: $y_n = Bx_n + \nu_n$

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

- ϵ_n and ν_n are random variables

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

Justified by the law of large numbers. . .

and the central limit theorem.

- Can also be viewed as approximating $\pi(dx) = f(x)dx$ with

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

Justified by Glivenko-Cantelli type results.

The Importance–Sampling Identity

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

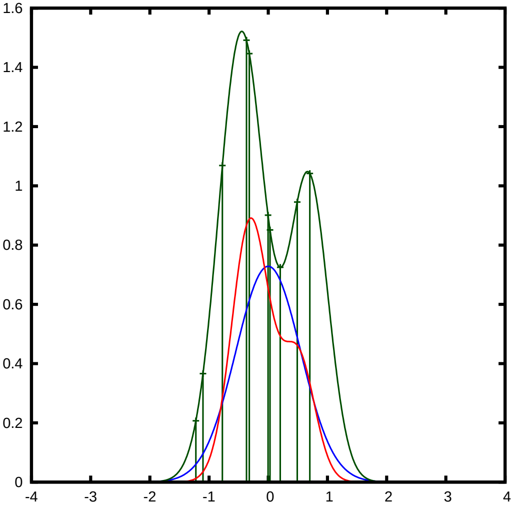
  define $w(x) = f(x)/g(x)$ and:

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

- 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:

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

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

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 $X_n^{(i)} \overset{\text{i.i.d.}}{\sim} q_n(\cdot|X_{1:n-1}^{(i)})$ or even $q_n(\cdot|X_{n-1}^{(i)})$.

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.

Importance Sampling on Huge Spaces Doesn't Work

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

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

- 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

- We can produce unweighted samples from weighted ones.
- Given $\{W_i, X_i\}_{i=1}^N$ a consistent resampling $\{\tilde{X}_i\}_{i=1}^N$ is such that

$$\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\varphi(\tilde{X}_i)\middle|\{W_i, X_i\}_{i=1}^N\right] = \sum_{i=1}^{N} W_i\varphi(X_i)$$

  for any continuous bounded φ.
- Simplest option: sample from empirical distribution

$$\tilde{X}_1, \ldots, \tilde{X}_N \overset{iid}{\sim} \sum_{j=1}^{N} W_j \delta_{X_j}(\cdot)$$

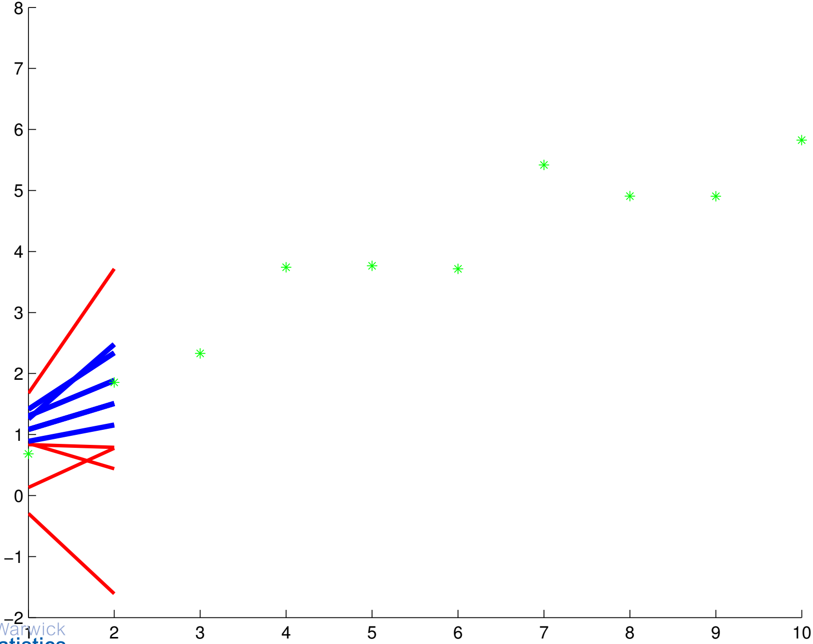
- Other approaches reduce the *additional* variance.

The SIR[esampling] Algorithm

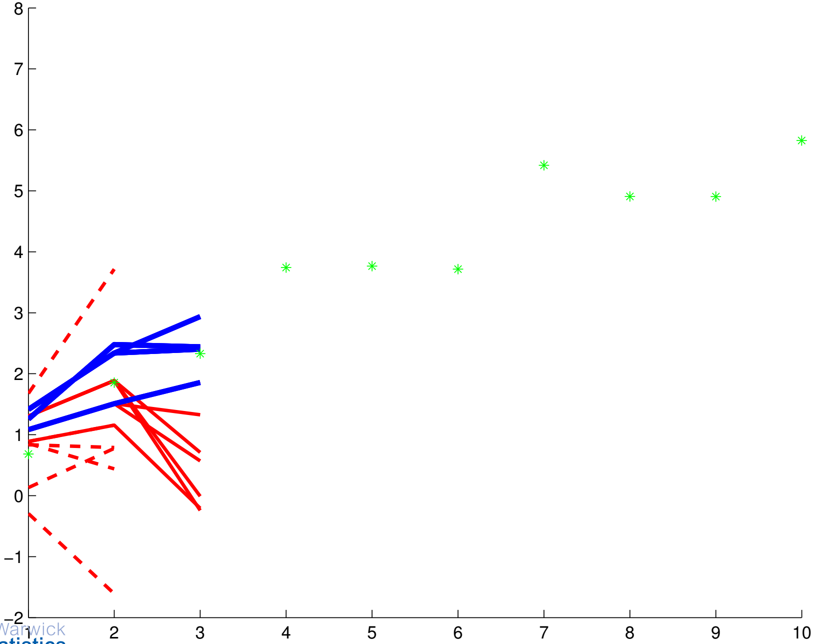
- Problem: variance of the weights builds up over time.
- Solution? Given $\{W_{n-1}^{(i)}, X_{1:n-1}^{(i)}\}$:
  1. **Resample**, to obtain $\{\frac{1}{N}, \widetilde{X}_{1:n-1}^{(i)}\}$.
  2. Sample $X_n^{(i)} \sim q_n(\cdot|\widetilde{X}_{n-1}^{(i)})$.
  3. Set $X_{1:n-1}^{(i)} = \widetilde{X}_{1:n-1}^{(i)}$.
  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...

  at least for problems like filtering.

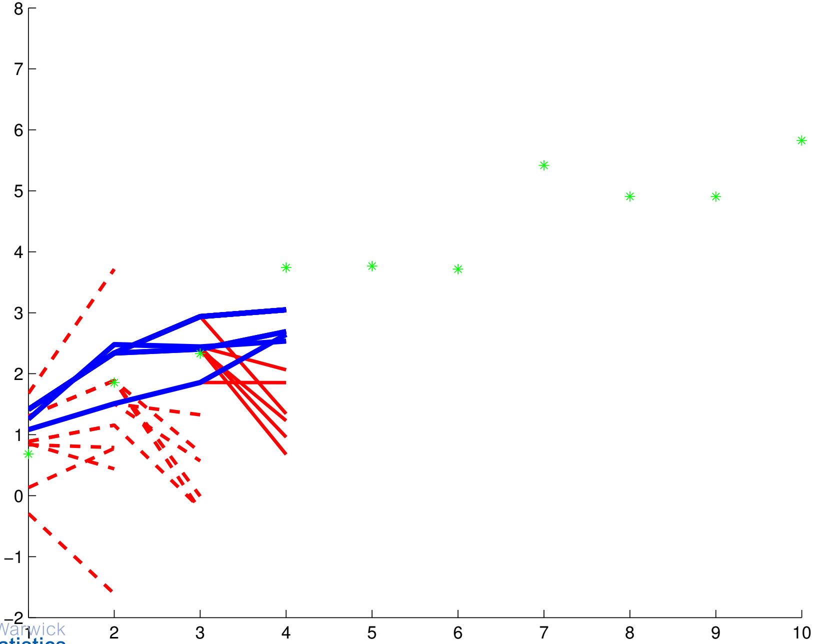
Iteration 2



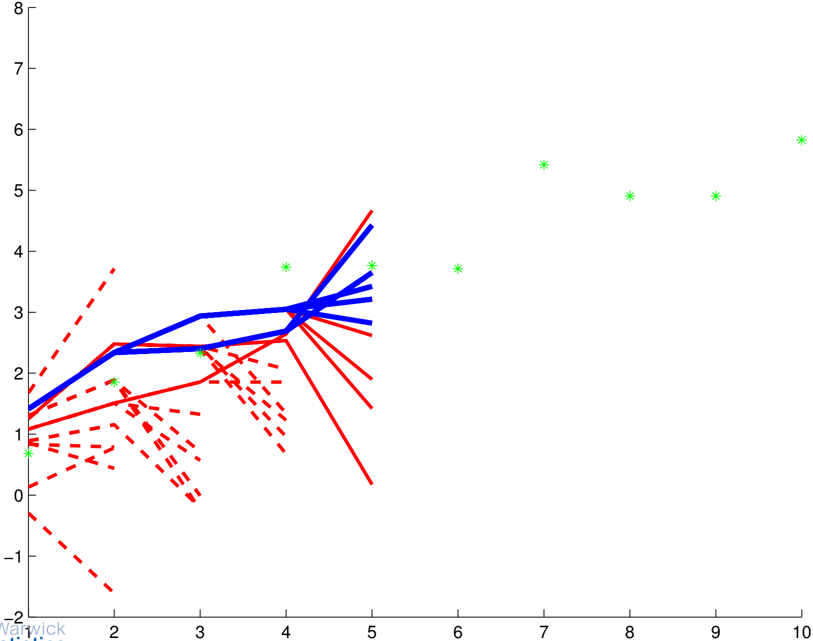
Iteration 3



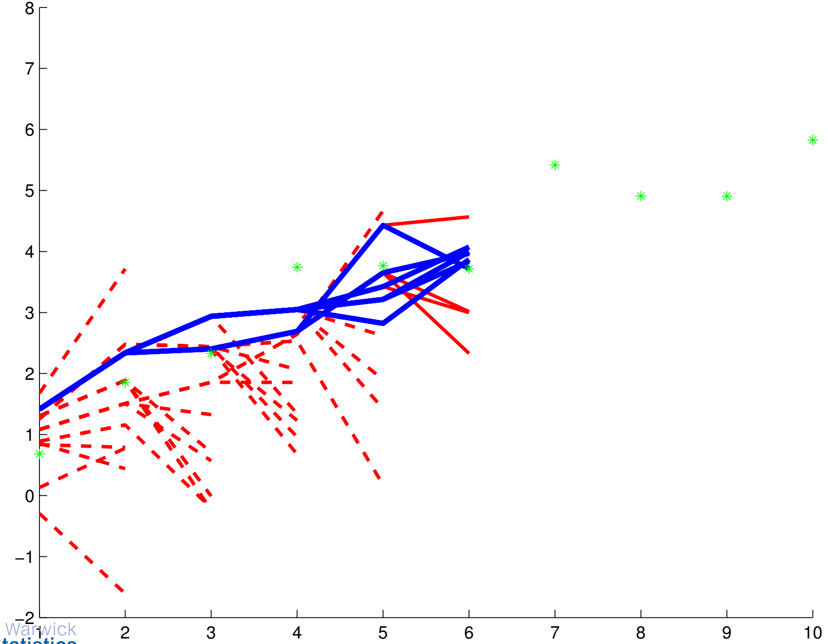
Iteration 4



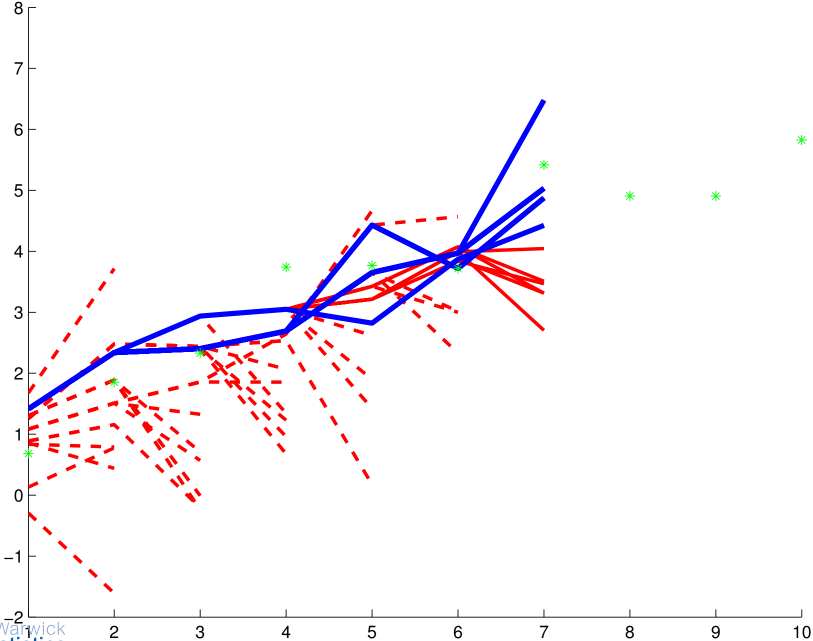
Iteration 5



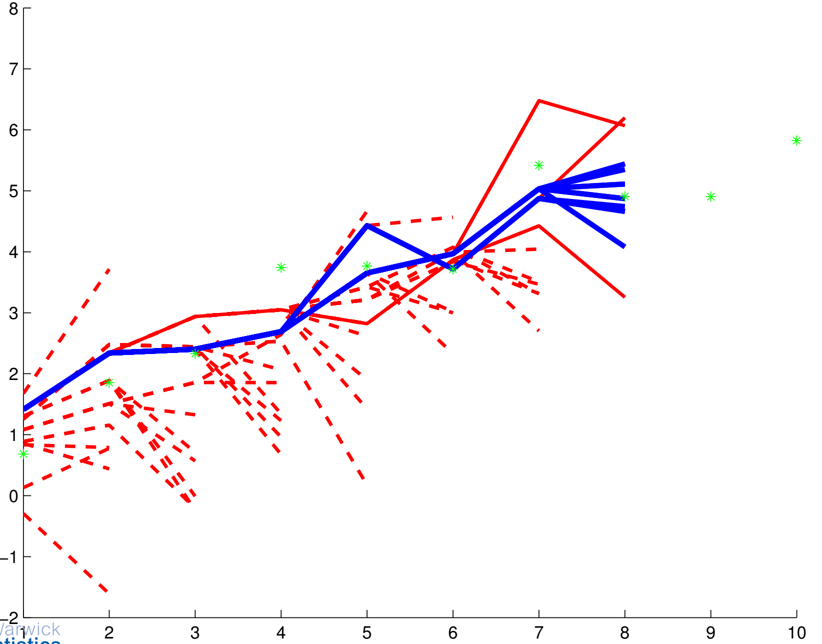
Iteration 6



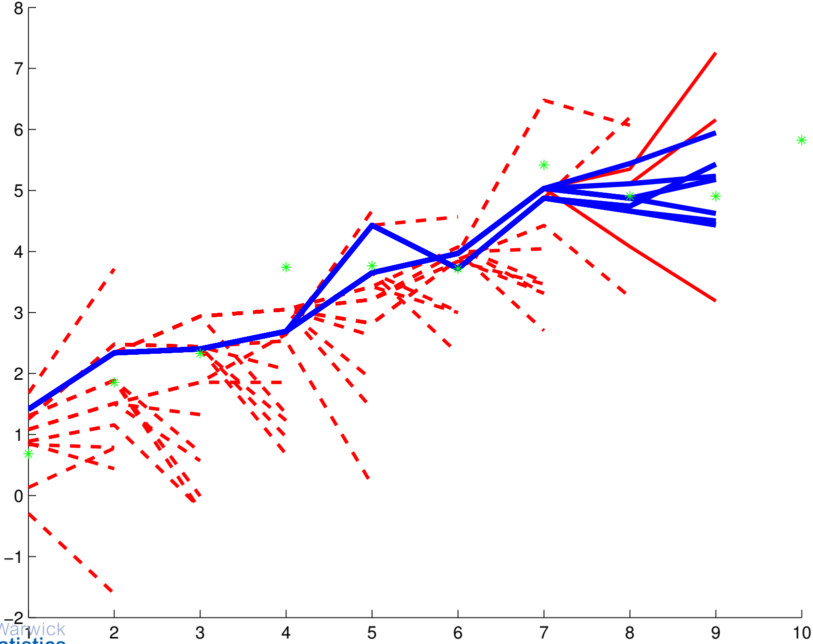
Iteration 7



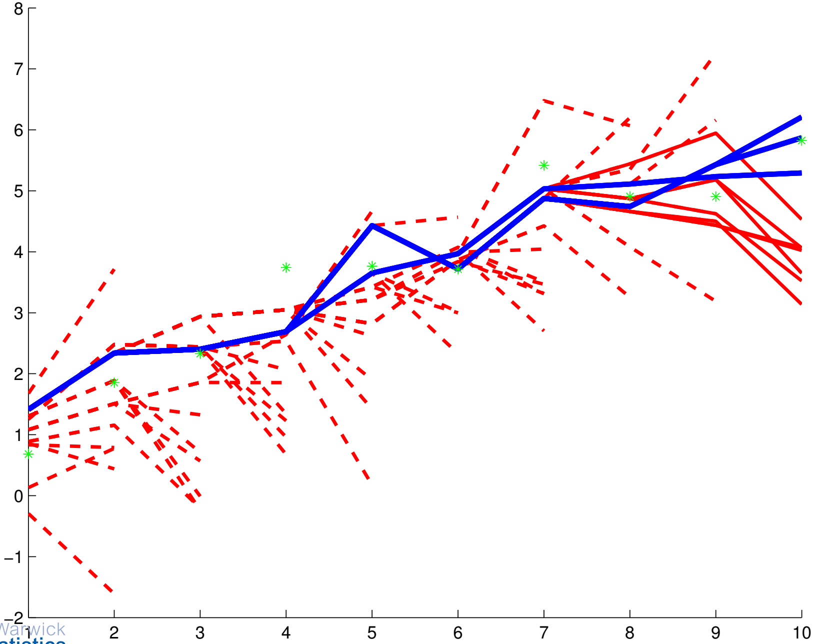
Iteration 8



Iteration 9



Iteration 10



Feynman-Kac Formulæ

A Probabilistic Perspective

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.

Elements of this framework:

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

Probabilistic Construction

Following Del Moral (2004)

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, $g, h : E \to \mathbb{R}$, define $g \cdot h : E \to \mathbb{R}$ via $(g \cdot h)(x) = g(x)h(x)$.

Given $e : E \to \mathbb{R}$ and $f : F \to \mathbb{R}$, let $(e \otimes f)(x, y) := e(x)f(y)$.

The Feynman-Kac Formulæ

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

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

- 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:

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

Feynman-Kac Marginal Measures

We are typically interested in marginals:

"Predicted"

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

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

"Updated"

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

$$\widehat{\eta}_n = \frac{\mathbb{E}\left[\varphi_n(X_n)\prod_{i=1}^{n} G_i(X_i)\right]}{\mathbb{E}\left[\prod_{i=1}^{n} G_i(X_i)\right]} = \widehat{\gamma}_n(\varphi_n)/\widehat{\gamma}_n(\mathbf{1})$$

Key property:

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

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

A Glimpse of the Theory

A Dynamic Systems View:
How do the marginal distributions evolve?

Don't worry about the details in these slides.

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:

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

- So:

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

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:

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

with

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

- So: $\Psi_{g(y|\cdot)}$: Prior $\to$ 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:

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

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

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

  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:*
  $\forall p \leq q : \gamma_q = \gamma_p Q_{p:q}$.

A Normalised Feynman-Kac Semigroupoid

- We previously established:

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

- 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: $\forall p \leq q : \ \eta_q = \eta_p \Phi_{p:q}$.

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:

$$\begin{array}{ccccc} \eta_{n-1} \in \mathcal{P}(E_{n-1}) & \xrightarrow{S_{n-1,\eta_{n-1}}} & \widehat{\eta}_n \in \mathcal{P}(E_{n-1}) & \xrightarrow{M_n} & \eta_n \in \mathcal{P}(E_n) \\ \xi_{n-1}^{(N)} \in E_{n-1}^N & \xrightarrow{\text{Select}} & \widehat{\xi}_n^{(N)} \in E_{n-1}^N & \xrightarrow{\text{Mutate}} & \xi_n^{(N)} \in E_n^N \end{array}$$

- 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 (conditionally independent):

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

- 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

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

Formally: $\eta_n^N - \eta_n = \sum_{p=1}^{n} \Phi_{p,n}(\eta_p^N) - \Phi_{p,n}(\Phi_p(\eta_{p-1}^N))$

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

$$\sigma_n^2(\varphi_n) = \sum_{q=1}^{n} \eta_q \left[(\bar{Q}_{q,n}(\varphi_n - \eta_n(\varphi_n)))^2 \right]$$

where

$$\bar{Q}_{q,n}(\varphi_n)(x_q) = Q_{q,n}(\varphi_n)(x_q) / \eta_q Q_{q,n}(\mathbf{1}).$$

Particle Filters

A Simple Application

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: $q(x_{n-1}, x_t) = f(x_n|x_{n-1})$
  - Weight: $w(x_n) \propto g(y_n|x_n)$
- Feynman-Kac model:
  - Mutation: $M_n(x_{n-1}, dx_n) = f(x_n|x_{n-1})dx_t$.
  - Potential: $G_n(x_n) = g(y_n|x_n)$.
- 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

$$\sqrt{N} \left(\frac{\sum_{i=1}^{N} W_n^{(i)} \varphi_n(X_n^{(i)})}{\sum_{j=1}^{N} W_n^{(j)}} - \int \varphi_n(x_n) p(x_n | y_{1:n}) dx_n \right) \xrightarrow{\text{d}} \mathcal{N}\left(0, \sigma^2_{BS,n}(\varphi_n)\right)$$

Bootstrap Particle Filter: Asymptotic Variance

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

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

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

One Novel Use of SMC

Solving Fredholm Equations of the First Kind

Joint work with Francesca R. Crucinio and Arnaud Doucet

A particle method for solving Fredholm integral equations of the first kind. ArXiv mathematics e-print 2009.09974, 2020.

Fredholm Equations of the First Kind

$$h(y) = \int g(y \mid x) f(x) \, dx$$

- applications in: density deconvolution, (medical) image processing, epidemiology, PDEs, nonlinear regression settings, ...
- inverse ill-posed problem

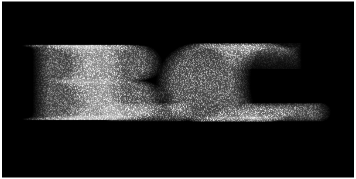
Applications I

$$h(y) = \int g(y - x) f(x) \, dx$$

- density deconvolution: recover the density of X from noisy observations $Y = X + \epsilon$
- epidemiology: recover incidence curve from observed death/hospitalization counts

Applications II

$$h(y) = \int g(x \mid y) f(x) \; dx$$



constant speed horizontal motion

Why Monte Carlo methods?

Standard Techniques

- require discretisation of the domain and/or make strong assumptions on f
- require discretisation of h
- impractical as dimension increases

Monte Carlo

- f, h probability densities, g density of a Markov kernel
- naturally implemented when we only have samples from h
- standard Monte Carlo rate $N^{-1/2}$

Regularisation

Approximating the solution of

$$h(y) = \int g(y \mid x) f(x)\ dx$$

by

$$f^\star = \text{argmin}\ \ \text{KL}(h, fg) := \int h(y) \log\left(\frac{h(y)}{\int g(y \mid x) f(x)\ dx}\right)\ dy$$

corresponds to maximum likelihood estimation.

Expectation Maximisation[1]

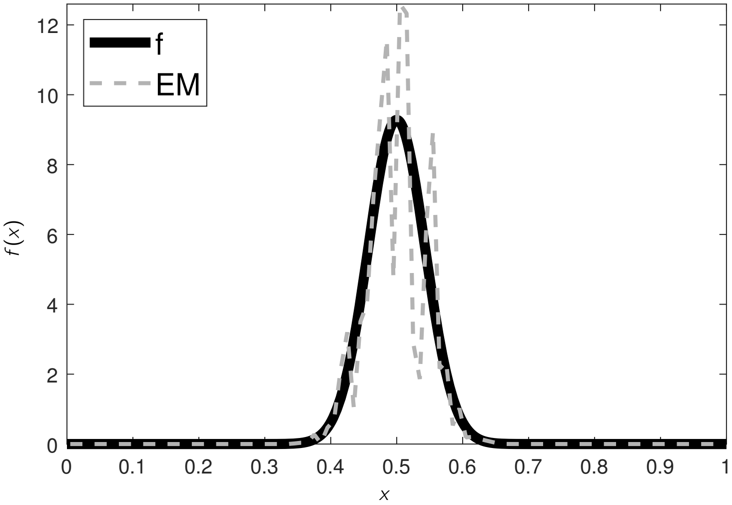
$$f_{n+1}(x) = f_n(x) \int \frac{h(y)g(y \mid x)}{\int g(y \mid z) f_n(z) dz} dy$$

and its brute-force discretisation

$$f_i^{(n+1)} = f_i^{(n)} \sum_{j=1}^{B_h} \left(\frac{h_j g_{ij}}{\sum_{k=1}^{B_f} g_{kj} f_k^{(n)}} \right) \qquad i = 1, \ldots, B_f .$$

[1] Kondor, Method of convergent weights, Nuclear Instruments and Methods in Physics Research (1983)

Inconsistency of MLE



Warwick Statistics

Expectation Maximisation Smoothing[2]

Obtain smoother reconstructions by addition of a smoothing step

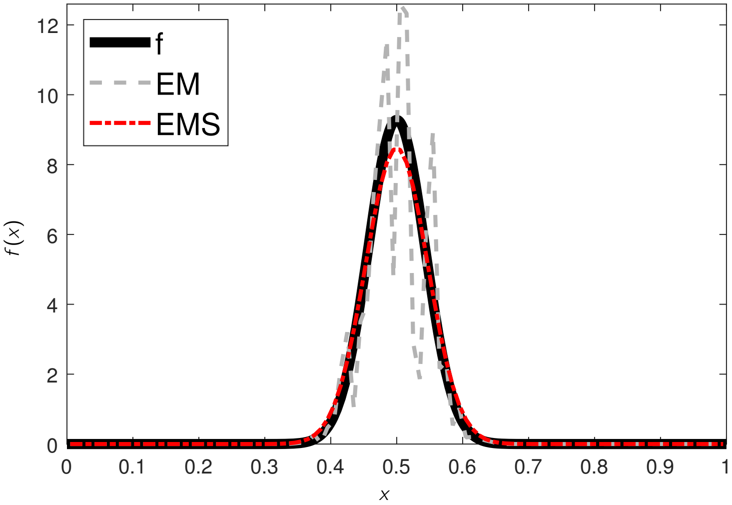
$$f_{n+1}(x) = \int K(x', x) f_n(x') \int \frac{g(y \mid x')}{\int g(y \mid z) f_n(z)\, dz} h(y)\, dy\, dx',$$

and its brute-force discretisation

$$\mathbf{f}^{(n+1)} = \mathsf{K}\mathbf{f}^{(n+1)}.$$

[2] Silverman et al., A smoothed EM approach to indirect estimation problems, JRSSB (1990)

Expectation Maximisation Smoothing (EMS)



EMS - Issues

1. Require discretisation of the domain of f
2. Assume f is piecewise constant
3. Impractical as dimension increases
4. What if we only have samples from h?

Recall Key Points of SMC

- a class of Monte Carlo methods that sequentially approximate a sequence of target probability densities $\{\eta_n(x_{1:n})\}$ of increasing dimension[3]
- the evolution of the sequence is described by alternating reweighting and proposing a new state

$$\eta_{n+1}(x_{1:n+1}) \propto \eta_n(x_{1:n}) G_n(x_n) M_{n+1}(x_{n+1} \mid x_n)$$

[3] We called these $\mathbb{Q}_n$ earlier...

SMC-EMS

1. Approximate f through a population of weighted samples
2. Provides an adaptive stochastic discretisation of the EMS recursion
3. Naturally deals with samples from h

SMC-EMS

Connect EMS with SMC:

$$f_{n+1}(x_{n+1}) = \int K(x_n, x_{n+1}) f_n(x_n) \int \frac{g(y_n \mid x_n)}{\int g(y_n \mid z) f_n(z)\ dz} h(y_n)\ dy_n\ dx_n,$$

Augment the density $\eta_n(x, y) = f_n(x)h(y)$

$$\eta_{n+1}(x_{n+1}, y_{n+1}) = \int \int \eta_n(x_n, y_n) K(x_n, x_{n+1}) \frac{g(y_n \mid x_n) h(y_{n+1})}{\int g(y_n \mid z) f_n(z)\ dz}\ dy_n\ dx_n$$

Remove the integral

$$\eta_{n+1}(x_{1:n+1}, y_{1:n+1}) = \eta_n(x_{1:n}, y_{1:n}) K(x_n, x_{n+1}) h(y_{n+1}) \frac{g(y_n \mid x_n)}{\int g(y_n \mid z)\ \eta_n(z) dz}$$

SMC-EMS — Idealised Version

Evolve a population of particles using the Markov kernels

$$M_{n+1}\left((x_{n+1}, y_{n+1}) \mid (x_n, y_n)\right) = K(x_n, x_{n+1})h(y_{n+1}),$$

the weight functions

$$G_n(x_n, y_n) = \frac{g(y_n \mid x_n)}{\int g(y_n \mid z)\eta_n(z)\, dz}$$

and a resampling mechanism.

Obtain an approximation f_{n+1}^N of f_{n+1} using kernel density estimation + smoothing with K.

SMC-EMS — Practical Version

Evolve a population of particles using the Markov kernels

$$M_{n+1}\left((x_{n+1}, y_{n+1}) \mid (x_n, y_n)\right) = K(x_n, x_{n+1})h(y_{n+1}),$$

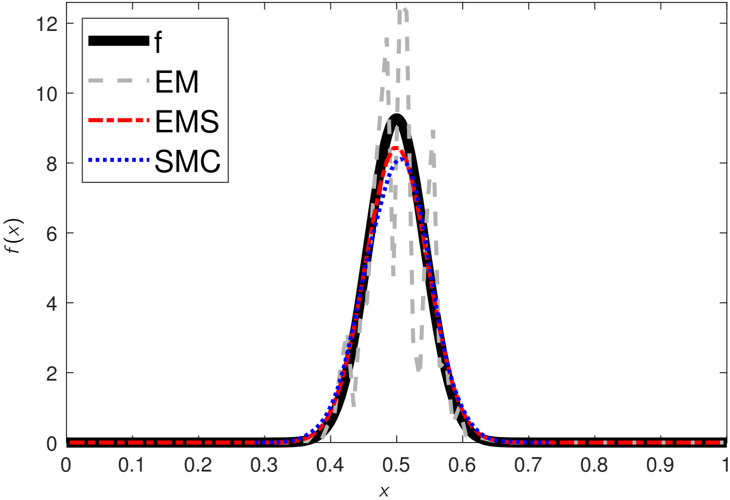
the weight functions

$$G_n^N(x_n, y_n) = \frac{g(y_n \mid x_n)}{\int g(y_n \mid z)\eta_n^N(dz)}$$

and a resampling mechanism.

Obtain an approximation f_{n+1}^N of f_{n+1} using kernel density estimation + smoothing with K.

It works!

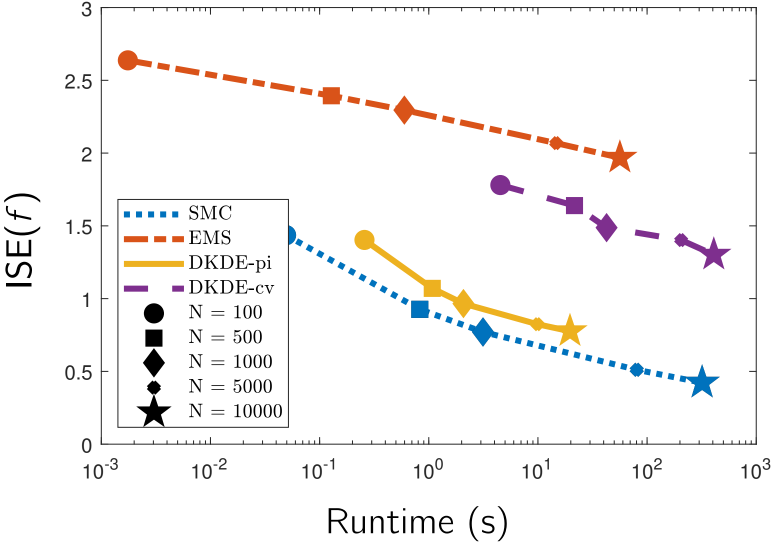


Density Deconvolution

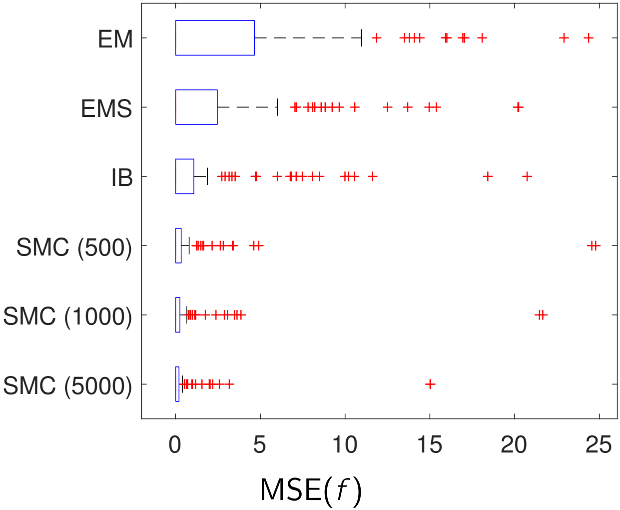
$$h(y) = \int g(y-x)f(x)\ dx$$

- Recover the density of X from noisy observations $Y = X + \epsilon$
- Estimators with optimal convergence rate exist: deconvolution kernel density estimators (DKDE)

Density Deconvolution — Results



Density Deconvolution — Smoothness



Motion Deblurring

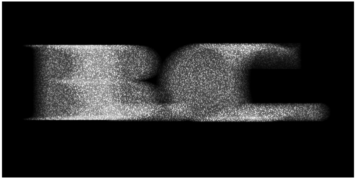
Model constant speed motion in the horizontal direction with

$$g(y_1, y_2 \mid x_1, x_2) = \mathcal{N}(y_2; x_2, \sigma^2)\text{Uniform}_{[-b/2, b/2]}(x_1 - y_1)$$

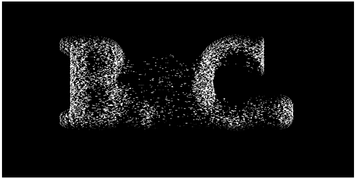
where b is the velocity and σ^2 is small.

Motion Deblurring

(a) Original image



(b) Blurred image + 0.5% noise



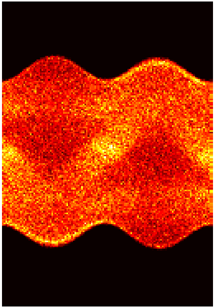
(c) RL

(d) SMC

Positron Emission Tomography

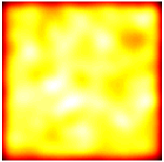
The reconstruction of a cross-section of the brain (a) from the data image provided by PET scanners (b) is described by a 2D Fredholm integral equation of the first kind.

(a) 128-pixels Shepp-Logan phantom

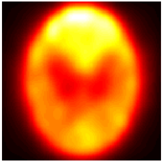


(b) Sinogram + noise

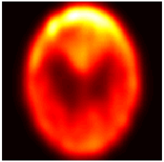
Positron Emission Tomography



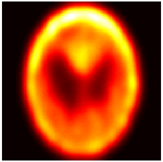
Iteration 1



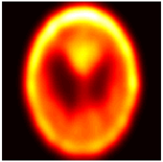
Iteration 5



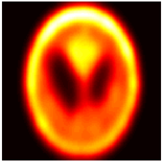
Iteration 10



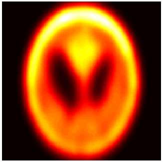
Iteration 15



Iteration 20



Iteration 50



Iteration 100

Reference Image

Effect of Dimension

Consider the simple toy example:

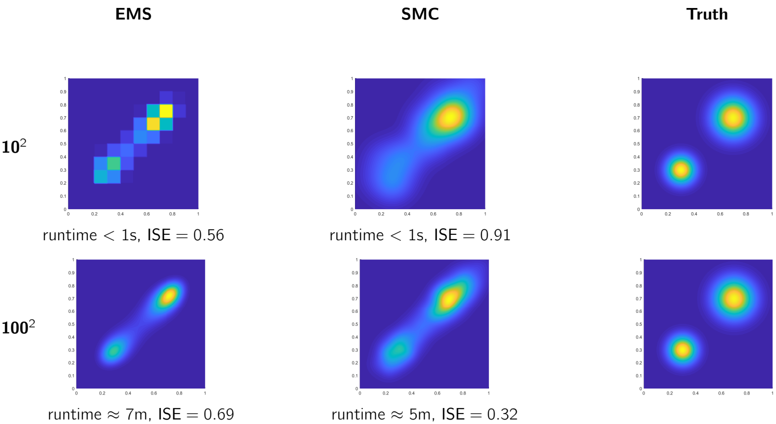
$$f(x_1, \ldots, x_{d_{\mathbb{X}}}) = \prod_{i=1}^{d_{\mathbb{X}}} \left(\frac{1}{3} \mathcal{N}(x_i; 0.3, 0.07^2) + \frac{2}{3} \mathcal{N}(x_i; 0.7, 0.1^2) \right),$$

$$g\left((y_1, \ldots, y_{d_{\mathbb{X}}}) | (x_1, \ldots, x_{d_{\mathbb{X}}})\right) = \prod_{i=1}^{d_{\mathbb{X}}} \mathcal{N}(y_i; x_i, 0.15^2),$$

$$h(y_1, \ldots, y_{d_{\mathbb{X}}}) = \prod_{i=1}^{d_{\mathbb{X}}} \left(\frac{1}{3} \mathcal{N}(x_i; 0.3, 0.07^2 + 0.15^2) + \frac{2}{3} \mathcal{N}(x_i; 0.7, 0.1^2 + 0.15^2) \right)$$

with $\mathbb{X} = \mathbb{Y} = \mathbb{R}^{d_{\mathbb{X}}}$.

Results



Reconstructions of a 2-dimensional mixture of Gaussian obtained with EMS and SMC. The number of bins/particles increases from 10^2 to 100^2. Runtime and accuracy are reported too.

	mean	variance	$\mathbb{P}(\square)$	$\mathbb{P}(\bigcirc)$	runtime $\log_{10}$s
$d_{\mathbb{X}} = 2$					
EMS - $B = 10$	**1.38e-04**	4.96e-05	5.30e-02	**7.04e-03**	-1.71
SMC - $N = 10^2$	3.87e-04	**1.26e-05**	**4.70e-02**	1.46e-02	**-2.02**
EMS - $B = 32$	1.42e-04	5.31e-05	5.17e-02	**5.86e-03**	1.28
SMC - $N = 10^3$	**4.29e-05**	**5.81e-06**	**3.02e-02**	8.29e-03	**0.94**
EMS - $B = 100$	1.42e-04	5.38e-05	5.15e-02	**6.11e-03**	5.31
SMC - $N = 10^4$	**3.84e-06**	**4.51e-06**	**2.77e-02**	8.57e-03	**5.11**
$d_{\mathbb{X}} = 3$					
EMS - $B = 5$	**2.53e-04**	1.26e-04	1.46e-01	8.59e-03	-1.47
SMC - $N = 10^2$	3.76e-04	**3.23e-05**	**7.41e-02**	**7.56e-03**	**-2.06**
EMS - $B = 10$	2.00e-04	5.75e-05	9.00e-02	2.42e-03	1.40
SMC - $N = 10^3$	**4.62e-05**	**8.50e-06**	**7.00e-02**	**1.54e-03**	**1.08**
EMS - $B = 22$	2.04e-04	6.12e-05	8.83e-02	1.64e-03	5.66
SMC - $N = 10^4$	**3.53e-06**	**6.68e-06**	**6.61e-02**	**9.38e-04**	**5.30**

	mean	variance	$\mathbb{P}(\square)$	$\mathbb{P}(\bigcirc)$	runtime $\log_{10}$s
$d_{\mathbb{X}} = 4$					
EMS - $B = 4$	**1.98e-04**	**1.55e-05**	1.22e-01	**1.16e-03**	-0.65
SMC - $N = 10^2$	4.77e-04	9.77e-05	**6.85e-02**	5.48e-03	**-2.08**
EMS - $B = 6$	2.43e-04	4.02e-05	1.09e-01	7.80e-04	1.70
SMC - $N = 10^3$	**3.45e-05**	**1.80e-05**	**8.68e-02**	**7.21e-04**	**0.95**
EMS - $B = 10$	2.60e-04	6.59e-05	1.03e-01	5.54e-04	5.32
SMC - $N = 10^4$	**4.10e-06**	**8.59e-06**	**8.95e-02**	**2.22e-04**	**5.12**
$d_{\mathbb{X}} = 5$					
EMS - $B = 3$	**5.66e-05**	2.67e-04	2.12e-01	**1.27e-02**	-0.56
SMC - $N = 10^2$	6.59e-04	**1.34e-04**	**3.89e-02**	1.41e-02	**-1.96**
EMS - $B = 4$	2.42e-04	**2.08e-05**	1.29e-01	**7.59e-04**	1.51
SMC - $N = 10^3$	**5.57e-05**	4.54e-05	**7.49e-02**	9.10e-04	**1.14**
EMS - $B = 7$	2.82e-04	5.71e-05	1.36e-01	2.09e-04	6.63
SMC - $N = 10^4$	**3.39e-06**	**1.27e-05**	**8.62e-02**	**5.73e-05**	**5.36**

Theoretical Guarantees

Assumptions

- f has compact support
- g is continuous, bounded above and below
- K has bounded density

The Monte Carlo algorithm has standard convergence rate $N^{-1/2}$

The estimator is well-founded:

- $\lim_{N\to\infty} \int |f_{n+1}^N(x) - f_{n+1}(x)| dx \overset{\text{a.s.}}{=} 0$
- $\lim_{N\to\infty} \mathbb{E}\left[\int |f_{n+1}^N(x) - f_{n+1}(x)|^2 \, dx\right] = 0$

Conclusions: SMC-EMS

- SMC-EMS is a novel method to solve Fredholm integral equations of the first kind based on a stochastic discretisation of EMS
- This method performs better that the state of the art on some toy and realistic examples
- The scheme inherits many convergence properties from SMC and KDE.

Summary

- SMC provides a mechanism for approximating (sequences of) probability distributions via importance sampling and resampling.
- There is still scope to further develop (and understand) SMC Methodology as the Fredholm equation example illustrates.
- My own current interests include:
  - Divide-and-conquer approaches to efficient distributed implementation.
  - The interaction with Generalized Bayesian Inference.
  - Automatic optimization of SMC algorithms.

Some References

- **Some approachable SMC References** A. Doucet and A. M. Johansen. A tutorial on particle filtering and smoothing: Fiteen years later. In D. Crisan and B. Rozovsky, editors, *The Oxford Handbook of Nonlinear Filtering*, pages 656–704. 2011.
  N. Chopin and O. Papaspiliopoulos. An introduction to Sequential Monte Carlo. Springer. 2020. See: SpringerLink.
- **SMC for Fredholm Equation of the 1st kind:**
  F.R. Crucinio, A. Doucet, and A.M. Johansen. A particle method for solving Fredholm integral equations of the first kind. ArXiv mathematics e-print 2009.09974, 2020.
- **MCMC for Fredholm Equations of the 2nd kind:**
  A. Doucet, A. M. Johansen, and V. B. Tadić. On solving integral equations using Markov Chain Monte Carlo. *Applied Mathematics and Computation*, 216:2869–2889, 2010.