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
19th January 2012
Part 1 – Statistics & Computation
Problems and Algorithms
A Statistical Problem
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:
  \[ Y_i | \{ X_i = x_i \} \sim g(\cdot | x_i). \]

- 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 $X_1, \ldots, X_N \sim \text{i.i.d. } f$.
  - Estimate $\hat{I} = \frac{1}{N} \sum_{i=1}^{N} \varphi(X_i)$.

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

$$\hat{\pi}^N(dx) = \frac{1}{N} \sum_{i=1}^{N} \delta_{X_i}(dx).$$
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 $X_1, \ldots, X_N \sim g$.
  - Estimate $\hat{I} = \frac{1}{N} \sum_{i=1}^{N} w(X_i)\varphi(X_i)$.

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

  $$
  \hat{\pi}^N(dx) = \frac{1}{N} \sum_{i=1}^{N} w(X_i)\delta_{X_i}(dx).
  $$
Importance Sampling Example
Self-Normalised Importance Sampling

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

\[
\frac{\mathbb{E}_g(\nu \varphi)}{\mathbb{E}_g(\nu 1)} = \frac{\mathbb{E}_g(cw \varphi)}{\mathbb{E}_g(cw 1)} = \frac{c \mathbb{E}_f(\varphi)}{c \mathbb{E}_f(1)} = \mathbb{E}_f(\varphi).
\]

- Estimate the numerator and denominator with the same sample:

\[
\hat{I} = \frac{\sum_{i=1}^{N} \nu(X_i) \varphi(X_i)}{\sum_{i=1}^{N} \nu(X_i)}.
\]

- Biased for finite samples, but consistent.
- Typically reduces variance.
Importance Sampling for Smoothing/Filtering

Sample \( \{X^{(i)}_{1:n}\} \) at time \( n \) from \( q_n(x_{1:n}) \), define

\[
w_n(x_{1:n}) \propto \frac{p(x_{1:n}|y_{1:n})}{q(x_{1:n})} = \frac{p(x_{1:n}, y_{1:n})}{q(x_{1:n})p(y_{1:n})} \sum_{m=2}^{n} f(x_{m}|x_{m-1})g(y_{m}|x_{m})
\]

set \( W_n^{(i)} = w_n(X^{(i)}_{1:n}) / \sum_{j} w_n(X^{(j)}_{1:n}) \),

then \( \{W_n^{(i)}, X^{(i)}_{n}\} \) is a consistently weighted sample.

This seems inefficient.
Sequential Importance Sampling (SIS) I

- **Importance weight**

\[
\begin{align*}
  w_n(x_{1:n}) & \propto \frac{f(x_1)g(y_1|x_1) \prod_{m=2}^{n} f(x_m|x_{m-1})g(y_m|x_m)}{q_n(x_{1:n})} \\
  & = \frac{f(x_1)g(y_1|x_1)}{q_n(x_1)} \prod_{m=2}^{n} \frac{f(x_m|x_{m-1})g(y_m|x_m)}{q_n(x_m|x_{1:m-1})}
\end{align*}
\]

- **Given** \( \{W_{n-1}^{(i)}, X_{1:n-1}^{(i)}\} \) targeting \( p(x_{1:n-1}|y_{1:n-1}) \)
  - Let \( q_n(x_{1:n-1}) = q_{n-1}(x_{1:n-1}) \),
  - sample \( X_{n}^{(i)} \) i.i.d. \( q_n(\cdot|X_{1:n-1}^{(i)}) \) or even \( q_n(\cdot|X_{n-1}^{(i)}) \).
Sequential Importance Sampling (SIS) II

And update the weights:

\[ w_n(x_{1:n}) = w_{n-1}(x_{1:n-1}) \frac{f(x_n|x_{n-1})g(y_n|x_n)}{q_n(x_n|x_{n-1})} \]

\[ W_n^{(i)} = w_n(X_{1:n}^{(i)}) \]

\[ = w_{n-1}(X_{1:n-1}^{(i)}) \frac{f(X_n^{(i)}|X_{n-1}^{(i)})g(y_n|X_n^{(i)})}{q_n(X_n^{(i)}|X_{n-1}^{(i)})} \]

\[ = W_{n-1}^{(i)} \frac{f(X_n^{(i)}|X_{n-1}^{(i)})g(y_n|X_n^{(i)})}{q_n(X_n^{(i)}|X_{n-1}^{(i)})} \]

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:
\[
\sqrt{N} \left[ \frac{1}{N} \sum_{i=1}^{N} w(X_i)\varphi(X_i) - \int \varphi(x)f(x)dx \right] \to \mathcal{N}(0, \text{Var}_g [w\varphi])
\]

▶ This is true.
▶ But it’s not enough.
▶ \text{Var}_g [w\varphi] 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 \( \{ \frac{1}{N}, \tilde{X}_{1:n-1}^{(i)} \} \).
  2. Sample \( X_n^{(i)} \sim q_n(\cdot | \tilde{X}_{n-1}^{(i)}) \).
  3. Set \( X_{1:n}^{(i)} = \tilde{X}_{1:n}^{(i)} \).
  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.

★ There are many algorithms for doing this...
Iteration 2
Iteration 3
Iteration 4
Iteration 8

The diagram shows a plot with axes labeled from -2 to 8 on both the x-axis and y-axis. There are multiple lines and markers on the graph, indicating a data series plotted over iterations. The exact nature of the data and the significance of the markers are not specified in the image.
Part 2 – Applied Probability
Feynman-Kac Formulae
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

Following Del Moral (2004)
The Canonical Markov Chain

- Consider the filtered probability space:
  $$(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n \in \mathbb{N}}, P_\mu)$$

- Let $\{X_n\}_{n \in \mathbb{N}}$ be a Markov chain such that for any $n \in \mathbb{N}$:
  $$P_\mu(X_{1:n} \in dx_{1:n}) = \mu(dx_1) \prod_{i=2}^{n} M_i(x_{i-1}, dx_i)$$

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

- $(E_i, \mathcal{E}_i)$ 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 \(\mu\) on \((E, \mathcal{E})\) and a Markov kernel, \(K : E \to \mathcal{P}(F)\), define:

\[
\mu(\varphi_E) := \int \mu(dx) \varphi_E(x)
\]

\[
\mu K(\varphi_F) := \int \mu(dx) K(x, dy) \varphi_F(y)
\]

\[
K(\varphi_F)(x) := \int K(x, dy) \varphi_F(y)
\]

with \(\varphi_E, \varphi_F\) 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:

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

Define two path measures weakly:

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

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

where \( \varphi_{1:n} : \bigotimes_{i=1}^{n} E_i \to \mathbb{R} \).
Example (Filtering via FK Formulae: 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:

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

$$= \mathbb{E} \left[ \varphi_{1:n}(X_{1:n}) \prod_{i=1}^{n-1} g(y_i|X_i) \right] / \mathbb{E} \left[ \prod_{i=1}^{n-1} g(y_i|X_i) \right]$$

$$= \int \left[ f(x_1) \prod_{i=2}^{n} f(x_i|x_{i-1}) \right] \left[ \prod_{j=1}^{n-1} g(y_j|x_j) \right] \varphi_{1:n}(x_{1:n}) dx_{1:n}$$

$$= \int \left[ f(x_1) \prod_{i=2}^{n} f(x_i|x_{i-1}) \right] \left[ \prod_{j=1}^{n-1} g(y_j|x_j) \right] dx_{1:n}$$

$$= \int p(x_{1:n}|y_{1:n-1}) \varphi_{1:n}(x_{1:n}) dx_{1:n}$$
Example (Filtering via FK Formulæ: Update/Filtering)

- Whilst:

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

\[
= \mathbb{E} \left[ \varphi_{1:n}(X_{1:n}) \prod_{i=1}^{n} g(y_i|X_i) \right] / \mathbb{E} \left[ \prod_{i=1}^{n} g(y_i|X_i) \right]
\]

\[
= \int \left[ f(x_1) \prod_{i=2}^{n} f(x_i|x_{i-1}) \right] \left[ \prod_{j=1}^{n} g(y_j|x_j) \right] \varphi_{1:n}(x_{1:n}) \, dx_{1:n}
\]

\[
= \int \frac{\left[ f(x_1) \prod_{i=2}^{n} f(x_i|x_{i-1}) \right] \left[ \prod_{j=1}^{n} g(y_j|x_j) \right]}{\int \left[ f(x_1) \prod_{i=2}^{n} f(x_i|x_{i-1}) \right] \left[ \prod_{j=1}^{n} g(y_j|x_j) \right] \, dx_{1:n}} \, dx_{1:n}
\]

\[
= \int p(x_{1:n}|y_{1:n}) \varphi_{1:n}(x_{1:n}) \, dx_{1:n}
\]
Feynman-Kac Marginal Measures

We are typically interested in marginals:

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

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

\[
\eta_n(\varphi_n) = \mathbb{Q}_n(\mathbf{1}_{1:n-1} \otimes \varphi_n)
\]

\[
\hat{\eta}_n = \hat{\mathbb{Q}}_n(\mathbf{1}_{1:n-1} \otimes \varphi_n)
\]

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

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

Key property:

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

\[
\hat{\eta}_n(A_n) = \int_{E_1 \times \ldots \times E_{n-1} \times A_n} \hat{\mathbb{Q}}_n(dx_{1:n})
\]
Analysis: Semigroup Structure

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

- The unnormalized marginals obey:
  \[
  \hat{\gamma}_n(\varphi_n) = \gamma_n(\varphi_n \cdot G_n) \quad \gamma_n(\varphi_n) = \hat{\gamma}_{n-1} M_n(\varphi_n)
  \]

- Whilst the normalized marginals satisfy:
  \[
  \hat{\eta}_n(\varphi_n) = \frac{\hat{\gamma}_n(\varphi_n)}{\hat{\gamma}_n(1)} = \frac{\gamma_n(\varphi_n \cdot G_n)}{\gamma_n(G_n)} = \frac{\eta_n(\varphi_n \cdot G_n)}{\eta_n(G_n)}
  \quad \eta_n(\varphi_n) = \frac{\gamma_n(\varphi_n)}{\gamma_n(1)} = \frac{\hat{\gamma}_{n-1} M_n(\varphi_n)}{\hat{\gamma}_{n-1} M_n(1)} = \frac{\hat{\eta}_{n-1} M_n(\varphi_n)}{\hat{\eta}_{n-1} M_n(1)} = \hat{\eta}_{n-1} M_n(\varphi_n)
  \]

- So:
  \[
  \hat{\eta}_n = \frac{\hat{\eta}_{n-1} M_n(\varphi_n \cdot G_n)}{\hat{\eta}_{n-1} M_n(G_n)}
  \]
The Boltzmann-Gibbs Operator

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

\[
\Psi_G : \mathcal{P}(E) \to \mathcal{P}(E)
\]

\[
\Psi_G : \nu \to \Psi_G(\nu)
\]

- The **Boltzmann-Gibbs** Operator \( \Psi_G \) is defined weakly by:

\[
\forall \varphi \in \mathcal{C}_b : \quad \Psi_G(\nu)(\varphi) = \frac{\nu(G \cdot \varphi)}{\nu(G)}
\]

- or equivalently, for all measurable sets \( A \):

\[
\Psi_G(A) = \frac{\nu(G \cdot \mathbb{1}_A)}{\nu(G)} = \frac{\int_A \nu(dx) G(x)}{\int_E \nu(dx') G(x')}
\]
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: $\Psi_{g(y|\cdot)} : \text{Prior} \rightarrow \text{Posterior}$.
Markov Semigroups

- A **semigroup** $S$ 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:

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

- A **semigroupoid**, \( S' \) comprises:
  - A set, \( S \).
  - A *partial* associative binary operation, \( \cdot \).

- A Markov Chain with inhomogeneous transitions \( M_n \) has a **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:

\[
\mathbb{P}(X_{n+m} \in A | X_m = x) = M_{m,n+m}(x, A).
\]
An Unnormalized Feynman-Kac Semigroupoid

We previously established:

\[ \gamma_n = \gamma_{n-1} M_n \quad \text{and} \quad \gamma_n(\varphi_n) = \gamma_n(\varphi_n \cdot G_n) \]

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:

\[ \forall p \leq q : \gamma_q = \gamma_p Q_{p:q} \]
A Normalised Feynman-Kac Semigroupoid

- We previously established:

\[ \eta_n = \hat{\eta}_{n-1} M_n(\varphi_n) \]

\[ \hat{\eta}_n = \frac{\eta_n(\varphi_n \cdot G_n)}{\eta_n(G_n)} \]

- From the definition of \( \Psi_{G_n} \): \( \hat{\eta}_n = \Psi_{G_n}(\eta_n) \).

- Defining \( \Phi_n : \mathcal{P}(E_{n-1}) \to \mathcal{P}(E_n) \) as:

\[ \Phi_n : \eta_{n-1} \rightarrow \Psi_{G_{n-1}}(\eta_{n-1}) M_n \]

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.
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 Formulae 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:

$$
S_{n-1}^{(N)}(\xi_{n-1}^{(N)}, d\xi_n^{(N)}) = \prod_{i=1}^{N} S_{n,\eta_{n-1}^{(N)}}(\xi_{n-1}^{(N,p)}, d\xi_n^{(N,p)})
$$

$$
M_n^{(N)}(\xi_{n-1}^{(N)}, d\xi_n^{(N)}) = \prod_{i=1}^{N} M_n(\xi_{n-1}^{(N,p)}, \xi_n^{(N,p)})
$$

it is clear that:

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

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

\[ \xi_{n-1}^{(N,i)} \sim \text{i.i.d.} \Psi G_{n-1}(\eta_{n-1}^{(N)}) \]

- Mutation:

\[ \xi_n^{(N,i)} \sim \text{i.i.d.} M_n(\widehat{\xi}_n^{(N,i)}, d\xi_n^{(N,i)}) \]

- Semigroupoid

\[ \mathbb{P}^N(\xi_n^{(N)} \in d\xi_n^{(N)} | \xi_{n-1}^{(N)}) = \prod_{i=1}^{N} \Phi_n(\eta_{n-1}^{(N)})(d\xi_n^{(N,i)}) \]
Selected Results
Local Error Decomposition

\[ \eta_1 \rightarrow \eta_2 = \Phi_2(\eta_1) \rightarrow \eta_3 = \Phi_{1:3}(\eta_1) \rightarrow \ldots \rightarrow \Phi_{1:n}(\eta_1) \]

\[ \downarrow \]

\[ \eta_1^N \rightarrow \Phi_2(\eta_1^N) \rightarrow \Phi_{1:3}(\eta_1^N) \rightarrow \ldots \rightarrow \Phi_{1:n}(\eta_1^N) \]

\[ \downarrow \]

\[ \eta_2^N \rightarrow \Phi_3(\eta_2^N) \rightarrow \ldots \rightarrow \Phi_{2:n}(\eta_2^N) \]

\[ \downarrow \]

\[ \eta_3^N \rightarrow \ldots \rightarrow \Phi_{3:n}(\eta_3^N) \]

\[ \vdots \]

\[ \downarrow \]

\[ \eta_{n-1}^N \rightarrow \Phi_n(\eta_{n-1}^N) \]

\[ \downarrow \]

\[ \eta_n^N \]

Formally:

\[ \eta_n^N - \eta_n = \sum_{i=1}^{n} \Phi_{i,n}(\eta_i^N) - \Phi_{i,n}(\Phi_i(\eta_{i-1}^N)) \]
A Key Martingale

Proposition (Del Moral, 2004 Proposition 7.4.1)

For each $n \geq 0$, $\varphi_n \in C_b(E_n)$ define:

$$\Gamma_{:,n}^N(\varphi_n) : p \in \{1, \ldots, n\} \rightarrow \Gamma_{p,n}(\varphi_n)$$

$$\Gamma_{p,n}^N(\varphi_n) := \gamma_p^N(Q_p,n\varphi_n) - \gamma_p(Q_p,n\varphi_n)$$

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

$$\Gamma_{p,n}^N(\varphi_n) = \sum_{q=1}^{p} \gamma_q^N(1) \left[ \eta_q^N(Q_q,n\varphi_n) - \eta_{q-1}^N K_{q,\eta_q^N}(Q_q,n\varphi_n) \right]$$

$$\langle \Gamma_{:,n}^N \rangle_p = \sum_{q=1}^{p} \gamma_q^N(1)^2 \eta_{q-1}^N \left[ K_{q,\eta_{q-1}^N}(Q_q,n(\varphi_n)) - \eta_{q-1}^N K_{q,\eta_{q-1}^N} Q_q,n(\varphi_n) \right]^2$$
Normalizing the Unnormalized

\[ \eta_n^N(\varphi_n) - \eta_n(\varphi_n) = \frac{\gamma_n^N(\varphi_n)}{\gamma_n^N(1)} - \frac{\gamma_n(\varphi_n)}{\gamma_n(1)} \]

\[ = \frac{\gamma_n(1)}{\gamma_n^N(1)} \left[ \frac{\gamma_n^N(\varphi_n)}{\gamma_n(1)} - \frac{\gamma_n(\varphi_n)}{\gamma_n(1)} \times \frac{\gamma_n^N(1)}{\gamma_n(1)} \right] \]

\[ = \frac{\gamma_n(1)}{\gamma_n^N(1)} \left[ \frac{\gamma_n^N(\varphi_n)}{\gamma_n(1)} - \eta_n(\varphi_n) \times \frac{\gamma_n^N(1)}{\gamma_n(1)} \right] \]

\[ = \frac{\gamma_n(1)}{\gamma_n^N(1)} \left[ \gamma_n \left( \frac{\varphi_n - \eta_n(\varphi_n)}{\gamma_n(1)} \right) \right] \]
Theorem (Del Moral 2004: Theorem 7.4.4)

Under regularity conditions, for any $n \geq 1$, $p \geq 1$, $\varphi_n \in C_b(E_n)$:

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

By a Borel-Cantelli argument:

$$\lim_{N \to \infty} \eta_n^N(\varphi_n) \xrightarrow{a.s.} \eta_n(\varphi_n).$$
Central Limit Theorem

Proposition (Del Moral 2004: Proposition 9.4.2)

Under regularity conditions, for any \( n \geq 1 \):

\[
\sqrt{N} (\eta_n^N(\varphi_n) - \eta_n(\varphi_n)) \xrightarrow{d} \mathcal{N} (0, \sigma_n^2(\varphi_n))
\]

where

\[
\sigma_n^2(\varphi_n) = \sum_{q=1}^{n} \eta_{q-1} [K_{q,\eta_{q-1}}(Q_q,n(\varphi_n)) - K_{q,\eta_{q-1}}(Q_q,n(\varphi_n))]^2
\]
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 $\{\frac{1}{N}, \tilde{X}_{1:n-1}^{(i)}\}$.  \hspace{1cm} \text{Selection}
2. Sample $X_{n}^{(i)} \sim q_{n}(\cdot | \tilde{X}_{n-1}^{(i)})$.  \hspace{1cm} \text{Mutation}
3. Set $X_{1:n-1}^{(i)} = \tilde{X}_{1:n-1}^{(i)}$.
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

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

CLT

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

\[ \sigma_{BS,n}^2(\varphi_n) = \int \frac{p(x_1|y_1:n)^2}{p(x_1)} \left( \int \varphi_n(x_n)p(x_n|y_2:n, x_1)dx_n - \bar{\varphi}_n \right)^2 dx_1 \]

\[ + \sum_{k=2}^{t-1} \int \frac{p(x_k|y_1:n)^2}{p(x_k|y_{1:k-1})} \left( \int \varphi_n(x_n)p(x_n|y_{k+1:n}, x_k)dx_n - \bar{\varphi}_n \right)^2 dx_{1:k} \]

\[ + \int \frac{p(x_n|y_1:n)^2}{p(x_n|y_{1:n-1})} (\varphi_n(x_n) - \bar{\varphi}_n)^2 dx_n. \]

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 $\{\frac{1}{N}, \tilde{X}_{1:n-1}^{(i)}\}$. Selection
  2. Sample $X_{n}^{(i)} \sim q_n(\cdot | \tilde{X}_{n-1}^{(i)})$. Mutation
  3. Set $X_{1:n-1}^{(i)} = \tilde{X}_{1:n-1}^{(i)}$.
  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.
$\sigma^2_{SIR,n}(\varphi_n) = \int \frac{p(x_1|y_1:n)^2}{q_1(x_1)} \left( \int \varphi_n(x_n)p(x_n|y_2:n, x_1)dx_n - \bar{\varphi}_n \right)^2 dx_1$

$$+ \sum_{k=2}^{t-1} \int \frac{p(x_{1:k}|y_1:n)^2}{p(x_{1:k-1}|y_{1:k-1})q_k(x_k|x_{k-1})} \left( \int \varphi_n(x_n)p(x_n|y_{k+1:n}, x_k)dx_n - \bar{\varphi}_n \right)^2 dx_{1:k}$$

$$+ \int \frac{p(x_{1:n}|y_1:n)^2}{p(x_{1:n-1}|y_{1:n-1})q_n(x_n|x_{n-1})} (\varphi_n(x_n) - \bar{\varphi}_n)^2 dx_{1:n}.$$
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 \( \lambda_n^{(i)} \propto \hat{p}(y_n|X_{n-1}^{(i)}) \).
- Propose new states, from the mixture distribution

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

- Weight samples, correcting for the pre-weighting.

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

- Resample particle set.
Some Well Known Refinements

We can tidy things up a bit:

1. The auxiliary variable step is equivalent to multinominal resampling.

2. So, there’s no need to resample before the pre-weighting.

Now we have:

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

$$W_n^{(i)} \propto \frac{f(X_{n,n}^{(i)}|X_{n,n-1}^{(i)})g(X_{n,n}^{(i)}|y_n)}{\lambda_n^{(i)} q_n(X_{n,n}^{(i)}|X_{n,1:n-1}^{(i)})}$$
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 $\lambda_{n+1}^{(i)} \propto \hat{p}(y_{n+1} | X_{n}^{(i)})$.

An SIR algorithm targeting:

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

\[
\sqrt{N} \left( \bar{\phi}_{n,APF} - \bar{\phi}_n \right) \rightarrow \mathcal{N} \left( 0, \sigma_n^2(\varphi_n) \right)
\]

where,

\[
\sigma_n^2(\varphi_n) = \\
\int \frac{p(x_1|y_{1:n})^2}{q_1(x_1)} \left( \int \varphi_n(x_n)p(x_n|y_{2:n}, x_1)dx_n - \bar{\varphi}_n \right)^2 dx_1 \\
+ \sum_{k=2}^{t-1} \int \frac{p(x_{1:k}|y_{1:n})^2}{\hat{p}(x_{1:k-1}|y_{1:k})q_k(x_k|x_{k-1})} \left( \int \varphi_n(x_n)p(x_n|y_{k+1:n}, x_k)dx_n - \bar{\varphi}_n \right)^2 dx_{1:k} \\
+ \int \frac{p(x_{1:n}|y_{1:n})^2}{\hat{p}(x_{1:n-1}|y_{1:n})q_n(x_n|x_{n-1})} (\varphi_n(x_n) - \bar{\varphi}_n)^2 dx_{1:n}.
\]
Practical Implications

- It means we’re doing importance sampling.
- Choosing $\hat{p}(y_n|x_{n-1}) = p(y_n|x_n = \mathbb{E}[X_n|x_{n-1}])$ is dangerous.
- A safer choice would be ensure that
  \[ \sup_{x_{n-1}, x_n} \frac{g(y_n|x_n)f(x_n|x_{n-1})}{\hat{p}(y_n|x_{n-1})q(x_n|x_{n-1})} < \infty \]
- Using APF doesn’t ensure superior performance.
A Contrived Illustration

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

\[ X = \{0, 1\} \quad p(x_1 = 0) = 0.5 \quad p(x_n = x_{n-1}) = 1 - \delta \]

\[ Y = X \quad p(y_n = x_n) = 1 - \varepsilon. \]

- \( \delta \) controls ergodicity of the state process.
- \( \varepsilon \) controls the information contained in observations.

Consider estimating \( \mathbb{E}(X_2 | Y_{1:2} = (0, 1)) \).
Variance of SIR - Variance of APF
Variance Comparison at $\epsilon = 0.25$

- SISR
- APF
- SISR Asymptotic
- APF Asymptotic

Graph showing the variance comparison at $\epsilon = 0.25$ with different symbols and lines for SISR, APF, SISR Asymptotic, and APF Asymptotic.
Further Reading
Particle Filters / Sequential Monte Carlo


Resampling


Feynman-Kac Particle Models


Auxiliary Particle Filters