# Asymptotic Genealogies of Sequential Monte Carlo Algorithms

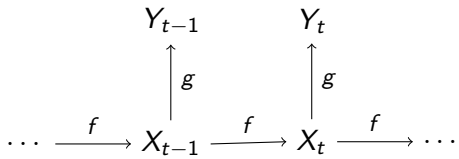
Suzie Brown

26 April 2019

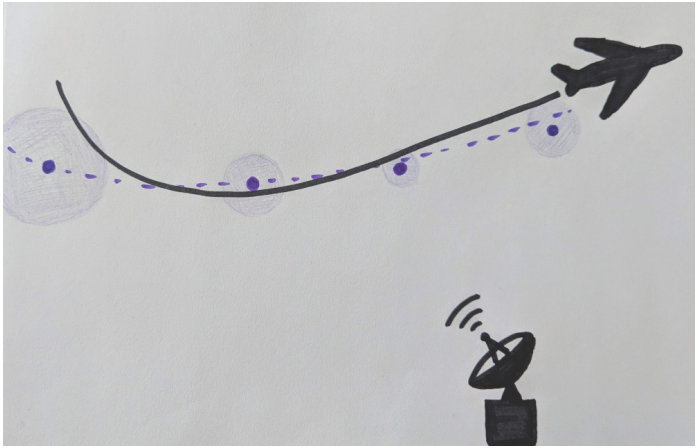
# State space models

$X_0, \ldots, X_T \in \mathcal{X}$
$Y_0, \ldots, Y_T \in \mathcal{Y}$

$X_0 \sim \mu(\cdot)$
$X_{t+1} \mid (X_t = x_t) \sim f(\cdot|x_t)$
$Y_t \mid (X_t = x_t) \sim g(\cdot|x_t)$



# Target tracking



use noisy radar data to infer position/trajectory of aircraft:

- $f$ models how aircraft moves
- $g$ models uncertainty in radar measurements

# Inference problems

**Filtering:** where is it now? $p(x_t|y_{0:t})$

**Prediction:** where will it go next? $p(x_{t+1}|y_{0:t})$

**Smoothing:** where has it been? $p(x_{0:t}|y_{0:t})$

Smoothing is "harder" than filtering/prediction.

# Deterministic solutions

## Kalman filter

Under a linear Gaussian model:

$$X_0 \sim \mathcal{N}(0, \Sigma_0)$$
$$X_{t+1} \mid (X_t = x_t) \sim \mathcal{N}(Ax_t, \Sigma_x)$$
$$Y_t \mid (X_t = x_t) \sim \mathcal{N}(Bx_t, \Sigma_y)$$

we can recursively compute filtering distributions. Then a backward pass of RTS smoother provides smoothing distributions.

This class of models is rather restrictive.

## Extended Kalman filter

In non-linear Gaussian models, use a local linear approximation and apply Kalman filter.

This requires gradients and performs poorly in models that are very non-linear.

## Unscented Kalman filter

For highly non-linear Gaussian models, replace the transition step of Kalman filter by propagating a representative set of points through $f$.

# Deterministic solutions

- Kalman filter provides optimal filter for linear Gaussian models, but extended/unscented Kalman filter are no longer optimal.
- All of these methods require model to be Gaussian. Deterministic solutions are available for some other conjugate families, but these are still restrictive.
- Solutions are also available in the case that $\mathcal{X}$ is finite (integrals become sums), but we will generally consider $\mathcal{X}$ to be a continuous space.

Sequential Monte Carlo (SMC) provides general-purpose (stochastic) methods that do not require a tractable model. Only requires sampling from $f(\cdot|x)$, and pointwise evaluation of $g(y|x)$ up to a normalising constant for each $y$.

# Sequential Monte Carlo

**Prior:** $p(x_{0:t}) = \mu(x_0) \prod_{i=1}^{t} f(x_i|x_{i-1})$

**Likelihood:** $p(y_{0:t}|x_{0:t}) = \prod_{i=0}^{t} g(y_i|x_i)$

**Posterior:** $p(x_{0:t}|y_{0:t}) \propto \mu(x_0) g(y_0|x_0) \prod_{i=1}^{t} f(x_i|x_{i-1}) g(y_i|x_i)$

- Represent posterior distribution at time $t$ with $N$ particles.
- Posterior factorises sequentially — avoid increase of dimension with $T$.

## Algorithm

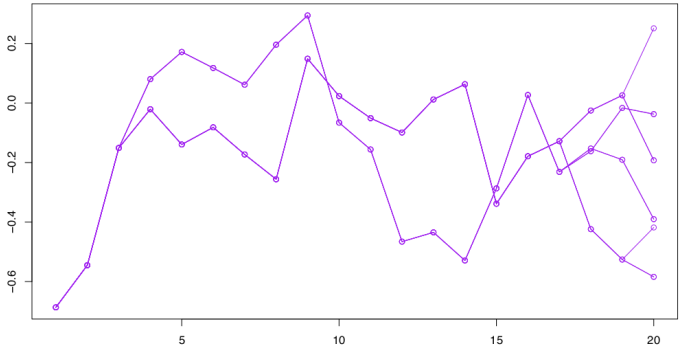
After initialisation, iterate these steps:

- **Propagate:** move the particles through the transition $f$
- **Calculate weights:** weight each particle according to $g$
- **Resample:** duplicate high-weight particles and kill off low-weight ones

Approximate posterior distribution $p(x_{0:t}|y_{0:t})$ by the empirical measure of the particles:

$$\hat{p}(x_{0:t}|y_{0:t}) = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_{0:t}^{(i)}}(x_{0:t})$$

# Ancestral degeneracy



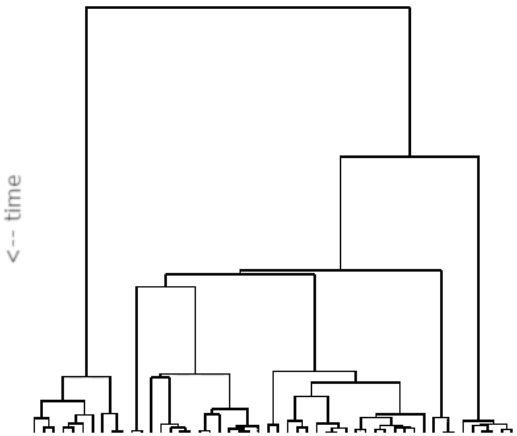

For smoothing we need a sample of trajectories.

Resampling means that trajectories of time $T$ particles coalesce backwards in time.

# Kingman's coalescent

Looking backwards in time, each pair of lineages coalesces with unit rate.

This is the limiting coalescent process in many population models as $N \to \infty$.



# Coalescent for SMC

- Analysing the coalescent can help us to understand the performance of various SMC algorithms.
- The limiting coalescent of an SMC algorithm may depend on the resampling mechanism.
- For multinomial resampling, the limiting coalescent is the Kingman coalescent.
- What about other (more used) resampling schemes?

# Resampling

- The total number of particles $N$ remains fixed.
- The particles after resampling are equally weighted.
- The resampling scheme is unbiased; that is, the expected number of offspring of each particle $i$ is equal to $Nw_t^{(i)}$.

# Resampling

**Multinomial:** $v_t^{(1:N)} \stackrel{d}{=} \text{Multinomial}(N, w_t^{(1:N)})$

**Residual:** $v_t^{(1:N)} \stackrel{d}{=} \lfloor N w_t^{(1:N)} \rfloor + \text{Multinomial}(R, r_t^{(1:N)}/R)$

$r_t^{(i)} := (N w_t^{(i)} - \lfloor N w_t^{(i)} \rfloor)$

$R := \sum r_t^{(i)}$

- Residual resampling yields lower Monte Carlo variance than multinomial resampling.
- Residual resampling is widely used by practitioners.
- Analysing the coalescent for residual resampling is a work in progress.

# Particle Gibbs

Hidden Markov model where transition depends on a parameter $\theta$:

$$\begin{aligned}
&\theta \sim \nu(\cdot) \\
&X_0 \sim \mu(\cdot) \\
&X_{t+1} \mid (X_t = x_t) \sim f_\theta(\cdot|x_t) \qquad t = 0, \ldots, T-1 \\
&Y_t \mid (X_t = x_t) \sim g(\cdot|x_t) \qquad t = 0, \ldots, T
\end{aligned}$$

**Gibbs sampler:** alternately sample from $p(\theta|x_{0:t}, y_{0:t})$ and $p(x_{0:t}|\theta, y_{0:t})$.

- SMC is an appropriate method for sampling from $p(x_{0:t}|\theta, y_{0:t})$
- To target the correct posterior distribution, need to use *conditional SMC*.