Imperial College London

Advanced Computational Methods in Statistics: Lecture 5 Sequential Monte Carlo/Particle Filtering

Axel Gandy

Department of Mathematics
Imperial College London
http://www2.imperial.ac.uk/~agandy

London Taught Course Centre
for PhD Students in the Mathematical Sciences
Autumn 2014

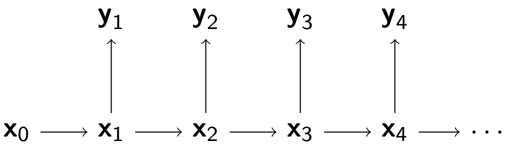
Outline

Sequential Monte Carlo/Particle Filtering

- Particle filtering introduced in Gordon et al. (1993)
- Most of the material on particle filtering is based on Doucet et al. (2001) and on the tutorial Doucet & Johansen (2008).
- Examples:
  - Tracking of Objects
  - Robot Localisation
  - Financial Applications

Setup - Hidden Markov Model

- $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots$: unobserved Markov chain - hidden states
- $\mathbf{y}_1, \mathbf{y}_2, \ldots$: observations;



- $\mathbf{y}_1, \mathbf{y}_2, \ldots$ are conditionally independent given $\mathbf{x}_0, \mathbf{x}_1, \ldots$
- Model given by
  - $\pi(\mathbf{x}_0)$ - the initial distribution
  - $f(\mathbf{x}_t|\mathbf{x}_{t-1})$ for $t \geq 1$ - the transition kernel of the Markov chain
  - $g(\mathbf{y}_t|\mathbf{x}_t)$ for $t \geq 1$ - the distribution of the observations
- Notation:
  - $\mathbf{x}_{0:t} = (\mathbf{x}_0, \ldots, \mathbf{x}_t)$ - hidden states up to time t
  - $\mathbf{y}_{1:t} = (\mathbf{y}_1, \ldots, \mathbf{y}_t)$ - observations up to time t
- Interested in the posterior distribution $p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})$ or in $p(\mathbf{x}_t|\mathbf{y}_{1:t})$

Some Remarks

- No explicit solution in the general case - only in special cases.
- Will focus on the time-homogeneous case, i.e. the transition densities f and the density of the observations g will be the same for each step.
  Extensions to inhomogeneous case straightforward.
- It is important that one is able to update quickly as new data becomes available, i.e. if $\mathbf{y}_{t+1}$ is observed want to be able to quickly compute $p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t+1})$ based on $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ and $\mathbf{y}_{t+1}$.

Example- Bearings Only Tracking

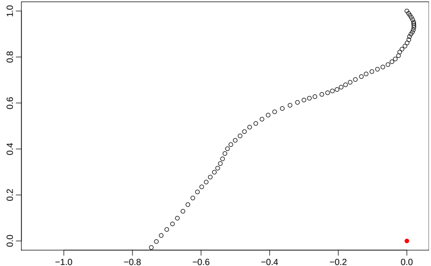
- Gordon et al. (1993); Ship moving in the two-dimensional plane
- Stationary observer sees only the angle to the ship.
- Hidden states: position $x_{t,1}, x_{t,2}$, speed $x_{t,3}, x_{t,4}$.
- Speed changes randomly

$$x_{t,3} \sim N(x_{t-1,3}, \sigma^2), \quad x_{t,4} \sim N(x_{t-1,4}, \sigma^2)$$

- Position changes accordingly

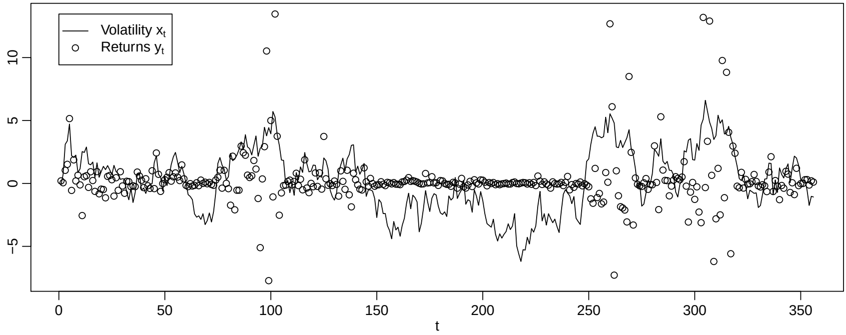
$$x_{t,1} = x_{t-1,1} + x_{t-1,3}, \quad x_{t,2} = x_{t-1,2} + x_{t-1,4}$$

- Observations: $y_t \sim N(\tan^{-1}(x_{t,1}/x_{t,2}), \eta^2)$



Example- Stochastic Volatility

- Returns: $y_t \sim N(0, \beta^2 \exp(x_t))$ (observable from price data)
- Volatility: $x_t \sim N(\alpha x_{t-1}, \frac{\sigma^2}{(1-\alpha)^2})$, $x_1 \sim N(0, \frac{\sigma^2}{(1-\alpha)^2})$,
- $\sigma = 1$, $\beta = 0.5$, $\alpha = 0.95$



Hidden Markov Chain with finitely many states

- If the hidden process $\mathbf{x}_t$ takes only finitely many values then $p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t+1})$ can be computed recursively via

$$p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t+1}) \propto g(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}) \sum_{\mathbf{x}_t} f(\mathbf{x}_{t+1}|\mathbf{x}_t) p(\mathbf{x}_t|y_{0:t})$$

- May not be practicable if there are too many states!

x_t can take 2 values only.

$P(X_0 = 1) = \pi_0(1)$

$P(X_0 = 2) = \pi_0(2)$

mult. with transition matrix

Effort: quadratic in # of states

$\mathcal{L}(X_1$

$g(k|y_1)$

$P(X_1 = k | Y_1 = y_1) = \dfrac{P(X_1 = k, Y_1 = y_1)}{P(Y_1 = y_1)}$

Kalman Filter

- Kalman (1960)
- Linear Gaussian state space model:
  - $\mathbf{x}_0 \sim N(\hat{\mathbf{x}}_0, P_0)$
  - $\mathbf{x}_t = A\mathbf{x}_{t-1} + \mathbf{w}_t$
  - $\mathbf{y}_t = H\mathbf{x}_t + \mathbf{v}_t$
  - $\mathbf{w}_t \sim N(0, Q)$ iid, $\mathbf{v}_t \sim N(0, R)$ iid
  - A, H deterministic matrices; $\hat{\mathbf{x}}_0$, P_0, A, H, Q, R known
- Explicit computation and updating of the posterior possible:

Posterior: $\mathbf{x}_t | \mathbf{y}_1, \ldots, \mathbf{y}_t \sim N(\hat{\mathbf{x}}_t, P_t)$

$\hat{\mathbf{x}}_0$, P_0

Time Update

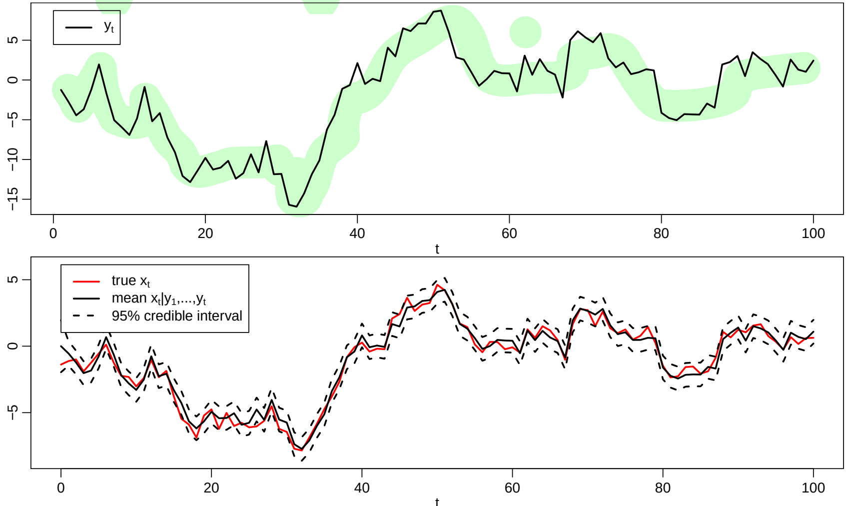
$$\hat{\mathbf{x}}_t^- = A\hat{\mathbf{x}}_{t-1}$$
$$P_t^- = AP_{t-1}A' + Q$$

Measurement Update

$$K_t = P_t^- H'(HP_t^- H' + R)^{-1}$$
$$\hat{\mathbf{x}}_t = \hat{\mathbf{x}}_t^- + K_t(\mathbf{y}_t - H\hat{\mathbf{x}}_t^-)$$
$$P_t = (I - K_t H)P_t^-$$

Kalman Filter - Simple Example

- $x_0 \sim N(0,1)$
- $x_t = 0.9x_{t-1} + w_t$
- $y_t = 2x_t + v_t$
- $w_t \sim N(0,1)$ iid, $v_t \sim N(0,1)$ iid



Kalman Filter - Remarks I

- Updating very easy - only involves linear algebra.
- Very widely used
- A, H, Q and R can change with time
- A linear control input can be incorporated, i.e. the hidden state can evolve according to

$$\mathbf{x}_t = A\mathbf{x}_{t-1} + B\mathbf{u}_{t-1} + \mathbf{w}_{t-1}$$

  where $\mathbf{u}_t$ can be controlled.
- Normal prior/updating/observations can be replaced through appropriate conjugate distributions.
- Continuous Time Version: Kalman-Bucy filter
  See Øksendal (2003) for a nice introduction.

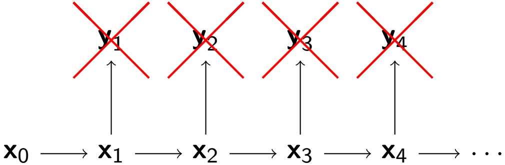
Kalman Filter - Remarks II

- Extended Kalman Filter
  extension to nonlinear dynamics:
  - $\mathbf{x}_t = f(\mathbf{x}_{t-1}, u_{t-1}, \mathbf{w}_{t-1})$
  - $\mathbf{y}_t = g(\mathbf{x}_t, \mathbf{v}_t)$.

  where f and g are nonlinear functions.
  The extended Kalman filter linearises the nonlinear dynamics around the current mean and covariance.
  To do so it uses the Jacobian matrices, i.e. the matrices of partial derivatives of f and g with respect to its components.
  The extended Kalman filter does no longer compute precise posterior distributions.

Outline

What to do if there were no observations...



Without observations $\mathbf{y}_i$ the following simple approach would work:

- Sample N particles following the initial distribution

$$\mathbf{x}_0^{(1)}, \ldots, \mathbf{x}_0^{(N)} \sim \pi(\mathbf{x}_0)$$

- For every step propagate each particle according to the transition kernel of the Markov chain:

$$\mathbf{x}_{i+1}^{(j)} \sim f(\cdot|\mathbf{x}_i^{(j)}), \quad j = 1, \ldots, N$$

- After each step there are N particles that approximate the distribution of $\mathbf{x}_i$.
- Note: very easy to update to the next step.

Importance Sampling

$x_{0:t}^{(i)} \sim p(x_{0:t} \mid y_{1:t}),\ i=1,\dots,n$

$\frac{1}{N}\sum_{i=1}^{N} \phi(x_{0:t}^{(i)})$

- Cannot sample from $\mathbf{x}_{0:t}|\mathbf{y}_{1:t}$ directly.
- Main idea: Change the density we are sampling from.
- Interested in $\mathrm{E}(\phi(\mathbf{x}_{0:t})|\mathbf{y}_{1:t}) = \int \phi(\mathbf{x}_{0:t}) p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) d\mathbf{x}_{0:t}$
- For any density h,

$$\mathrm{E}(\phi(\mathbf{x}_{0:t})|\mathbf{y}_{1:t}) = \int \phi(\mathbf{x}_{0:t}) \frac{p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})}{h(\mathbf{x}_{0:t})} h(\mathbf{x}_{0:t}) d\mathbf{x}_{0:t},$$

- Thus an unbiased estimator of $\mathrm{E}(\phi(\mathbf{x}_{0:t})|\mathbf{y}_{1:t})$ is

$$\hat{I} = \frac{1}{N}\sum_{i=1}^{N} \phi(\mathbf{x}_{0:t}^{i}) w_t^i,$$

where $w_t^i = \frac{p(\mathbf{x}_{0:t}^i|\mathbf{y}_{1:t})}{h(\mathbf{x}_{0:t}^i)}$ and $\mathbf{x}_{0:t}^1, \dots, \mathbf{x}_{0:t}^N \sim h$ iid.

- How to evaluate $p(\mathbf{x}_{0:t}^i|\mathbf{y}_{1:t})$?
- How to choose h? Can importance sampling be done recursively?

Sequential Importance Sampling I

- Recursive definition and sampling of the importance sampling distribution:

$$h(\mathbf{x}_{0:t}) = h(\mathbf{x}_t|\mathbf{x}_{0:t-1})h(\mathbf{x}_{0:t-1})$$

- Can the weights be computed recursively? By Bayes' Theorem:

$$w_t = \frac{p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})}{h(\mathbf{x}_{0:t})} = \frac{p(\mathbf{y}_{1:t}|\mathbf{x}_{0:t})p(\mathbf{x}_{0:t})}{h(\mathbf{x}_{0:t})p(\mathbf{y}_{1:t})}$$

Hence,

$$w_t = \frac{g(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{y}_{1:t-1}|\mathbf{x}_{0:t-1})f(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{0:t-1})}{h(\mathbf{x}_t|\mathbf{x}_{0:t-1})h(\mathbf{x}_{0:t-1})p(\mathbf{y}_{1:t})}$$

Thus,

$$w_t = w_{t-1}\frac{g(\mathbf{y}_t|\mathbf{x}_t)f(\mathbf{x}_t|\mathbf{x}_{t-1})}{h(\mathbf{x}_t|\mathbf{x}_{0:t-1})}\frac{p(\mathbf{y}_{1:t-1})}{p(\mathbf{y}_{1:t})}$$

Sequential Importance Sampling I

- Recursive definition and sampling of the importance sampling distribution:

$$h(\mathbf{x}_{0:t}) = h(\mathbf{x}_t|\mathbf{x}_{0:t-1})h(\mathbf{x}_{0:t-1})$$

- Can the weights be computed recursively? By Bayes' Theorem:

$$w_t = \frac{p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})}{h(\mathbf{x}_{0:t})} = \frac{p(\mathbf{y}_{1:t}|\mathbf{x}_{0:t})p(\mathbf{x}_{0:t})}{h(\mathbf{x}_{0:t})p(\mathbf{y}_{1:t})}$$

Hence,

$$w_t = \frac{g(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{y}_{1:t-1}|\mathbf{x}_{0:t-1})f(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{0:t-1})}{h(\mathbf{x}_t|\mathbf{x}_{0:t-1})h(\mathbf{x}_{0:t-1})p(\mathbf{y}_{1:t})}$$

Thus,

$$w_t = w_{t-1}\frac{g(\mathbf{y}_t|\mathbf{x}_t)f(\mathbf{x}_t|\mathbf{x}_{t-1})}{h(\mathbf{x}_t|\mathbf{x}_{0:t-1})}\frac{p(\mathbf{y}_{1:t-1})}{p(\mathbf{y}_{1:t})}$$

Sequential Importance Sampling I

- Recursive definition and sampling of the importance sampling distribution:

$$h(\mathbf{x}_{0:t}) = h(\mathbf{x}_t|\mathbf{x}_{0:t-1})h(\mathbf{x}_{0:t-1})$$

- Can the weights be computed recursively? By Bayes' Theorem:

$$w_t = \frac{p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})}{h(\mathbf{x}_{0:t})} = \frac{p(\mathbf{y}_{1:t}|\mathbf{x}_{0:t})p(\mathbf{x}_{0:t})}{h(\mathbf{x}_{0:t})p(\mathbf{y}_{1:t})}$$

Hence,

$$w_t = \frac{g(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{y}_{1:t-1}|\mathbf{x}_{0:t-1})f(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{0:t-1})}{h(\mathbf{x}_t|\mathbf{x}_{0:t-1})h(\mathbf{x}_{0:t-1})p(\mathbf{y}_{1:t})}$$

does not involve x_t

Thus,

$$w_t = w_{t-1}\frac{g(\mathbf{y}_t|\mathbf{x}_t)f(\mathbf{x}_t|\mathbf{x}_{t-1})}{h(\mathbf{x}_t|\mathbf{x}_{0:t-1})}\frac{p(\mathbf{y}_{1:t-1})}{p(\mathbf{y}_{1:t})}$$

Sequential Importance Sampling II

- Can work with normalised weights: $\tilde{w}_t^i = \frac{w_t^i}{\sum_j w_t^j}$; then one gets the recursion

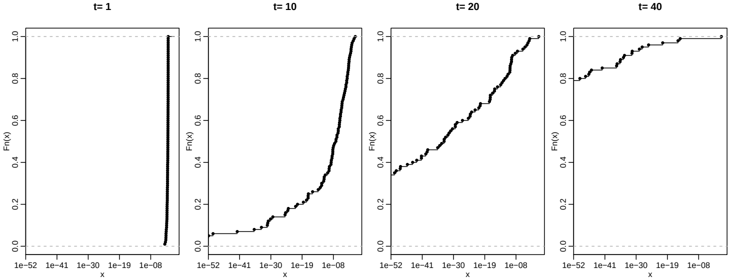
$$\tilde{w}_t^i \propto \tilde{w}_{t-1}^i \frac{g(\mathbf{y}_t|\mathbf{x}_t^i)f(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i)}{h(\mathbf{x}_t^i|\mathbf{x}_{0:t-1}^i)}$$

- If one uses uses the prior distribution $h(\mathbf{x}_0) = \pi(\mathbf{x}_0)$ and $h(\mathbf{x}_t|\mathbf{x}_{0:t-1}) = f(\mathbf{x}_t|\mathbf{x}_{t-1})$ as importance sampling distribution then the recursion is simply

$$\tilde{w}_t^i \propto \tilde{w}_{t-1}^i g(\mathbf{y}_t|\mathbf{x}_t^i)$$

Failure of Sequential Importance Sampling

- Weights degenerate as t increases.
- Example: $x_0 \sim N(0,1)$, $x_{t+1} \sim N(x_t, 1)$, $y_t \sim N(x_t, 1)$.
  - $N = 100$ particles
  - Plot of the empirical cdfs of the normalised weights $\tilde{w}_t^1, \dots, w_t^N$



  Note the log-scale on the x-axis.
  - Most weights get very small.

Resampling

- Goal: Eliminate particles with very low weights.
- Suppose

$$Q = \sum_{i=1}^{N} \tilde{w}_t^i \delta_{\mathbf{x}_t^i}$$

is the current approximation to the distribution of $\mathbf{x}_t$.
- Then one can obtain a new approximation as follows:
  - Sample N iid particles $\tilde{\mathbf{x}}_t^i$ from Q
  - The new approximation is

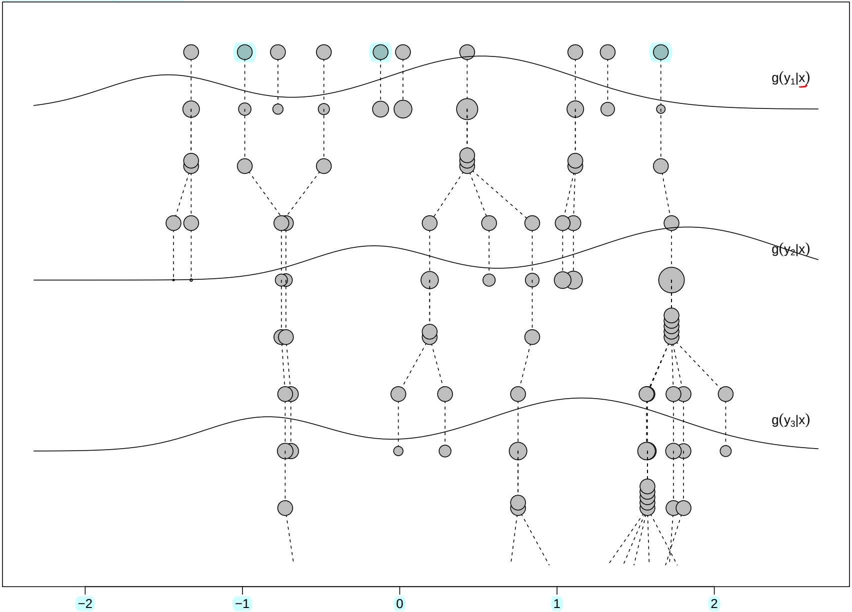
$$\tilde{Q} = \sum_{i=1}^{N} \frac{1}{N} \delta_{\tilde{\mathbf{x}}_t^i}$$

The Bootstrap Filter

1. Sample $\mathbf{x}_0^{(i)} \sim \pi(\mathbf{x}_0)$ and set $t = 1$
2. Importance Sampling Step
   For $i = 1, \ldots, N$:
   - Sample $\tilde{\mathbf{x}}_t^{(i)} \sim f(\mathbf{x}_t|\mathbf{x}_{t-1}^{(i)})$ and set $\tilde{\mathbf{x}}_{0:t}^{(i)} = (\mathbf{x}_{0:t-1}^{(i)}, \tilde{\mathbf{x}}_t^{(i)})$
   - Evaluate the importance weights $\tilde{w}_t^{(i)} = g(\mathbf{y}_t|\tilde{\mathbf{x}}_t^{(i)})$.
3. Selection Step:
   - Resample with replacement N particles $(\mathbf{x}_{0:t}^{(i)}; i = 1, \ldots, N)$ from the set $\{\tilde{\mathbf{x}}_{0:t}^{(1)}, \ldots, \tilde{\mathbf{x}}_{0:t}^{(1)}\}$ according to the normalised importance weights $\frac{\tilde{w}_t^{(i)}}{\sum_{j=1}^N \tilde{w}_t^{(j)}}$.
   - t:=t+1; go to step 2.

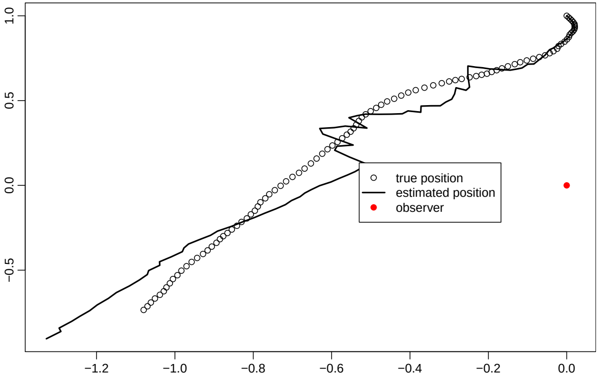
Illustration of the Bootstrap Filter

N=10 particles



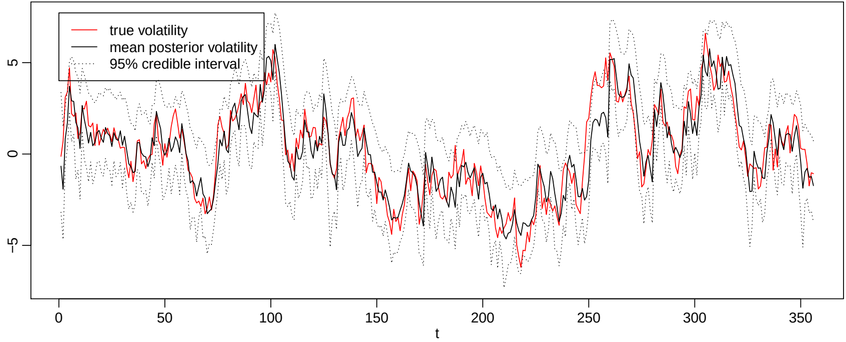
Example- Bearings Only Tracking

- $N = 10000$ particles



Example- Stochastic Volatility

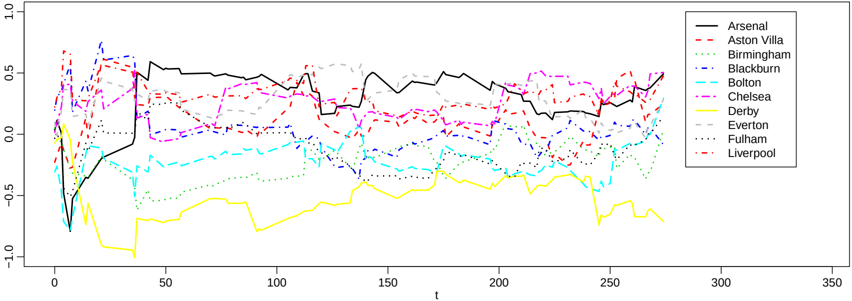
- Bootstrap Particle Filter with $N = 1000$

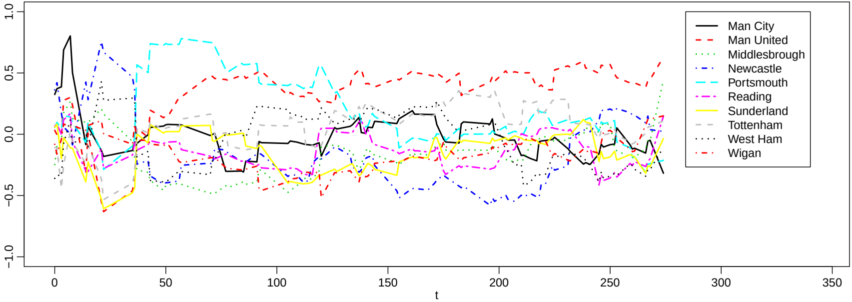


Example: Football

- Data: Premier League 2007/08
- $x_{t,j}$ "strength" of the jth team at time t, $j = 1, \dots, 20$
- $\mathbf{y}_t$ result of the games on date t
- Note: not time-homogeneous (different teams playing one-another - different time intervals between games).
- Model:
  - Initial distribution of the strength: $x_{t,j} \sim N(0, 1)$
  - Evolution of strength: $x_{t,j} \sim N((1 - \Delta/\beta)^{1/2} x_{t-\Delta,j}, \Delta/\beta)$ will use $\beta = 50$
  - Result of games conditional on strength: Match between team H of strength x_H (at home) against team A of strength x_A. Goals scored by the home team $\sim \mathit{Poisson}(\lambda_H \exp(x_H - x_A))$ Goals scored by the away team $\sim \mathit{Poisson}(\lambda_A \exp(x_A - x_H))$ λ_H and λ_A constants chosen based on the average number of goals scored at home/away.

Mean Team Strength



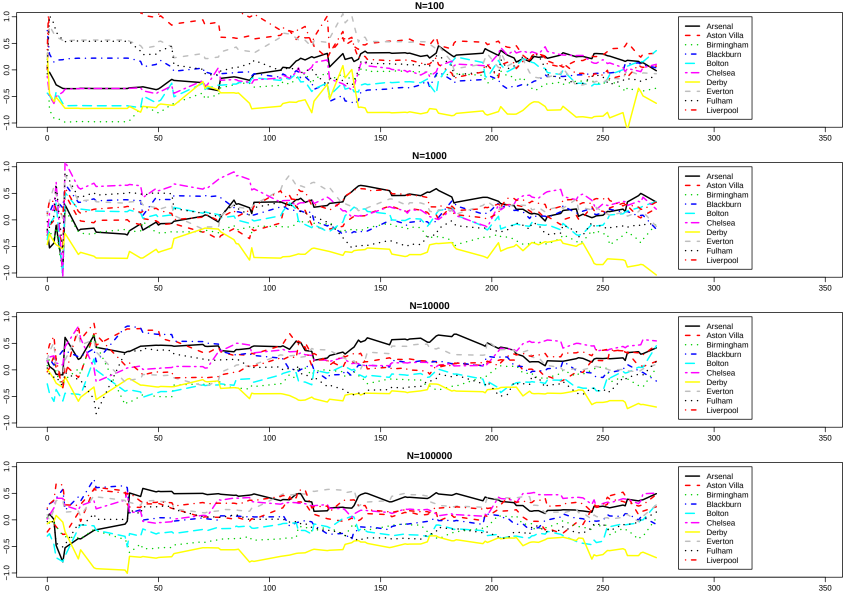


(based on $N = 100000$ particle)

League Table at the end of 2007/08

1	Man Utd	87
2	Chelsea	85
3	Arsenal	83
4	Liverpool	76
5	Everton	65
6	Aston Villa	60
7	Blackburn	58
8	Portsmouth	57
9	Manchester City	55
10	West Ham Utd	49
11	Tottenham	46
12	Newcastle	43
13	Middlesbrough	42
14	Wigan Athletic	40
15	Sunderland	39
16	Bolton	37
17	Fulham	36
18	Reading	36
19	Birmingham	35
20	Derby County	11

Influence of the Number N of Particles



Theoretical Results

- Convergence results are as $N \to \infty$
- Laws of Large Numbers
- Central limit theorems
  see e.g. Chopin (2004)
- The central limit theorems yield an asymptotic variance. This asymptotic variance can be used for theoretical comparisons of algorithms.

Outline

General Proposal Distribution

Algorithm with a general proposal distribution h:

1. Sample $\mathbf{x}_0^{(i)} \sim \pi(\mathbf{x}_0)$ and set $t = 1$
2. Importance Sampling Step
   For $i = 1, \ldots, N$:
   - Sample $\tilde{\mathbf{x}}_t^{(i)} \sim h_t(\mathbf{x}_t|\mathbf{x}_{t-1}^{(i)})$ and set $\tilde{\mathbf{x}}_{0:t}^{(i)} = (\mathbf{x}_{0:t-1}^{(i)}, \tilde{\mathbf{x}}_t^{(i)})$
   - Evaluate the importance weights $\tilde{w}_t^{(i)} = \frac{g(\mathbf{y}_t|\tilde{\mathbf{x}}_t^{(i)}) f(\tilde{\mathbf{x}}_t^{(i)}|\mathbf{x}_{t-1}^{(i)})}{h_t(\tilde{\mathbf{x}}_t^{(i)}|\mathbf{x}_{t-1}^{(i)})}$.
3. Selection Step:
   - Resample with replacement N particles $(\mathbf{x}_{0:t}^{(i)}; i = 1, \ldots, N)$ from the set $\{\tilde{\mathbf{x}}_{0:t}^{(1)}, \ldots, \tilde{\mathbf{x}}_{0:t}^{(1)}\}$ according to the normalised importance weights $\frac{\tilde{w}_t^{(i)}}{\sum_{j=1}^N \tilde{w}_t^{(j)}}$.
   - t:=t+1; go to step 2.

Optimal proposal distribution depends on quantity to be estimated. generic choice: choose proposal to minimise the variance of the normalised weights.

Improving Resampling

- Resampling was introduced to remove particles with low weights
- Downside: adds variance

Other Types of Resampling

- Goal: Reduce additional variance in the resampling step
- Standardised weights $W^1, \dots, W^N$;
- N^i - number of 'offspring' of the ith element.
- Need $\mathrm{E}\, N^i = W^i N$ for all i.
- Want to minimise the resulting variance of the weights.

Multinomial Resampling - resampling with replacement

Systematic Resampling
- Sample $U_1 \sim U(0, 1/N)$. Let $U_i = U_1 + \frac{i-1}{N}$, $i = 2, \dots, N$
- $N^i = |\{j : \sum_{k=1}^{i-1} W^k \leq U_j \leq \sum_{k=1}^{i} W^k|$

Residual Resampling
- Idea: Guarantee at least $\tilde{N}^i = \lfloor W^i N \rfloor$ offspring of the ith element;
- $\bar{N}^1, \dots, \bar{N}^n$: Multinomial sample of $N - \sum \tilde{N}^i$ items with weights $\bar{W}^i \propto W^i - \tilde{N}^i / N$.
- Set $N^i = \tilde{N}^i + \bar{N}^i$.

Adaptive Resampling

- Resampling was introduced to remove particles with low weights.
- However, resampling introduces additional randomness to the algorithm.
- Idea: Only resample when weights are "too uneven".
- Can be assessed by computing the variance of the weights and comparing it to a threshold.
- Equivalently, one can compute the "effective sample size" (ESS):

$$\mathrm{ESS} = \left(\sum_{i=1}^{n} (w_t^i)^2\right)^{-1}.$$

  ($w_t^1, \ldots, w_t^n$ are the normalised weights)
- Intuitively the effective sample size describes how many samples from the target distribution would be roughly equivalent to importance sampling with the weights w_t^i.
- Thus one could decide to resample only if

$$ESS < k$$

  where k can be chosen e.g. as $k = N/2$.

Outline

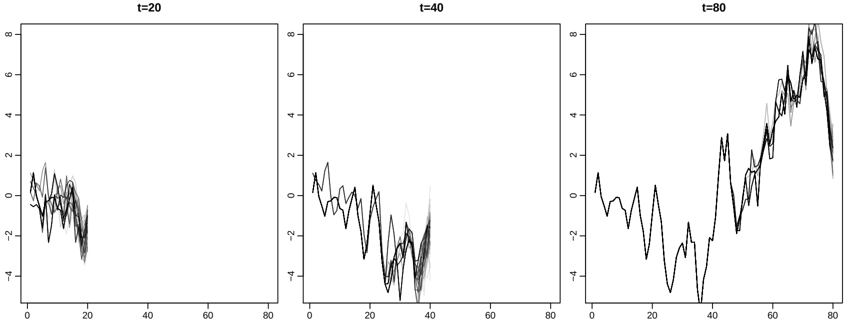
Path Degeneracy

Let $s \in \mathbb{N}$.

$$\#\{\mathbf{x}^i_{0:s} : i = 1, \dots, N\} \to 1 \quad (\text{as \# of steps } t \to infty)$$

Example

$x_0 \sim N(0,1)$, $x_{t+1} \sim N(x_t, 1)$, $y_t \sim N(x_t, 1)$. $N = 100$ particles.



Particle Smoothing

- Estimate the distribution of the state $\mathbf{x}_t$ given all the observations $\mathbf{y}_1, \dots, \mathbf{y}_\tau$ up to some late point $\tau > t$.
- Intuitively, a better estimation should be possible than with filtering (where only information up to $\tau = t$ is available).
- Trajectory estimates tend to be smoother than those obtained by filtering.
- More sophisticated algorithms are needed.

Filtering and Parameter Estimates

- Recall that the model is given by
  - $\pi(\mathbf{x}_0)$ - the initial distribution
  - $f(\mathbf{x}_t|\mathbf{x}_{t-1})$ for $t \geq 1$ - the transition kernel of the Markov chain
  - $g(\mathbf{y}_t|\mathbf{x}_t)$ for $t \geq 1$ - the distribution of the observations
- In practical applications these distributions will not be known explicitly - they will depend on unknown parameters themselves.
- Two different starting points
  - Bayesian point of view: parameters have some prior distribution
  - Frequentist point of view: parameters are unknown constants.
- Examples:

  Stoch. Volatility: $x_t \sim N(\alpha x_{t-1}, \frac{\sigma^2}{(1-\alpha)^2})$, $x_1 \sim N(0, \frac{\sigma^2}{(1-\alpha)^2})$,
  $y_t \sim N(0, \beta^2 \exp(x_t))$
  Unknown parameters: σ, β, α

  Football: Unknown Parameters: β, λ_H, λ_A.
- How to estimate these unknown parameters?

Maximum Likelihood Approach

- Let $\theta \in \Theta$ contain all unknown parameters in the model.
- Would need marginal density $p_\theta(y_{1:t})$.
- Can be estiamted by running the particle filter for each θ of interest and by muliplying the unnormalised weights, see the slides Sequential Importance Sampling I/II earlier in the lecture for some intuition.

Artificial(?) Random Walk Dynamics

- Allow the parameters to change with time - give them some dynamic.
- More precisely:
  - Suppose we have a parameter vector $\boldsymbol{\theta}$
  - Allow it to depend on time ($\boldsymbol{\theta}_t$),
  - assign a dynamic to it, i.e. a prior distribution and some transition probability from $\boldsymbol{\theta}_{t-1}$ to $\boldsymbol{\theta}_t$
  - incorporate $\boldsymbol{\theta}_t$ in the state vector $\mathbf{x}_t$
- May be reasonable in the Stoch. Volatility and Football Example

Bayesian Parameters

- Prior on $\boldsymbol{\theta}$
- Want: Posterior $p(\boldsymbol{\theta}|\mathbf{y}_1, \dots, \mathbf{y}_t)$.
- Naive Approach:
  - Incorporate $\boldsymbol{\theta}$ into $\mathbf{x}_t$; transition for these components is just the identity
  - Resampling will lead to $\boldsymbol{\theta}$ degenerating - after a moderate number of steps only few (or even one) $\boldsymbol{\theta}$ will be left
- New approaches: PMCMC
  - Particle filter within an MCMC algorithm, Andrieu et al. (2010) - computationally very expensive.
  - SMC^2 by Chopin, Jacob, Papaspiliopoulos, arXiv:1101.1528.

Now published

Outline

Concluding Remarks

- Active research area
- A collection of references/resources regarding SMC `http://www.stats.ox.ac.uk/~doucet/smc_resources.html`
- Collection of application-oriented articles: Doucet et al. (2001)
- Brief introduction to SMC: (Robert & Casella, 2004, Chapter 14)
- R-package implementing several methods: pomp `http://pomp.r-forge.r-project.org/`

Part I

Appendix

References I

Andrieu, C., Doucet, A. & Holenstein, R. (2010). Particle markov chain monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **72**, 269–342.

Chopin, N. (2004). Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statist.* **32**, 2385–2411.

Doucet, A., Freitas, N. D. & Gordon, N. (eds.) (2001). *Sequential Monte Carlo Methods in Practice*. Springer.

Doucet, A. & Johansen, A. M. (2008). A tutorial on particle filtering and smoothing: Fifteen years later. Available at `http://www.cs.ubc.ca/~arnaud/doucet_johansen_tutorialPF.pdf`.

Gordon, N., Salmond, D. & Smith, A. (1993). Novel approach to nonlinear/non-gaussian bayesian state estimation. *Radar and Signal Processing, IEE Proceedings F* **140**, 107–113.

Kalman, R. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering* **82**, 35–45.

References II

Øksendal, B. (2003). *Stochastic Differential Equations: An Introduction With Applications*. Springer.

Robert, C. & Casella, G. (2004). *Monte Carlo Statistical Methods*. Second ed., Springer.