Tutorial on Particle filters

Keith Copsey

Pattern and Information Processing Group

DERA Malvern

K.Copsey@signal.dera.gov.uk
Outline

- Introduction to particle filters
  - Recursive Bayesian estimation
- Bayesian Importance sampling
  - Sequential Importance sampling (SIS)
  - Sampling Importance resampling (SIR)
- Improvements to SIR
  - On-line Markov chain Monte Carlo
- Basic Particle Filter algorithm
- Examples
- Conclusions
- Demonstration
Particle Filters

- Sequential Monte Carlo methods for on-line learning within a Bayesian framework.

- Known as
  - Particle filters
  - Sequential sampling-importance resampling (SIR)
  - Bootstrap filters
  - Condensation trackers
  - Interacting particle approximations
  - Survival of the fittest
Recursive Bayesian estimation (I)

- Recursive filter:
  - System model:
    \[ x_k = f_k(x_{k-1}, \omega_k) \iff p(x_k | x_{k-1}) \]
  - Measurement model:
    \[ y_k = h_k(x_k, \nu_k) \iff p(y_k | x_k) \]
  - Information available:
    \[ D_k = (y_1, \ldots, y_k) \]
    \[ p(x_0) \]
Recursive Bayesian estimation (II)

- Seek: \( p(x_{k+i} | D_k) \)
  - \( i = 0 \): filtering.
  - \( i > 0 \): prediction.
  - \( i < 0 \): smoothing.

- Prediction:
  \[
  p(x_k | D_{k-1}) = \int p(x_k | x_{k-1}) p(x_{k-1} | D_{k-1}) \, dx_{k-1}
  \]
  since:
  \[
  p(x_k | D_{k-1}) = \int p(x_k, x_{k-1} | D_{k-1}) \, dx_{k-1}
  \]

NCAF January Meeting, Aston University, Birmingham.
Recursive Bayesian estimation (III)

- Update:

\[ p(x_k \mid D_k) = \frac{p(y_k \mid x_k) p(x_k \mid D_{k-1})}{p(y_k \mid D_{k-1})} \]

- where:

\[ p(y_k \mid D_{k-1}) = \int p(y_k \mid x_k) p(x_k \mid D_{k-1}) \, dx_k \]

- since:

\[ p(y_k \mid D_{k-1}) = \int p(y_k, x_k \mid D_{k-1}) \, dx_k \]
Classical approximations

- **Analytical methods:**
  - Extended Kalman filter,
  - Gaussian sums… (Alspach et al. 1971)
  
  • Perform poorly in numerous cases of interest

- **Numerical methods:**
  - point masses approximations,
  - splines. (Bucy 1971, de Figueiro 1974…)
  
  • Very complex to implement, not flexible.
Perfect Monte Carlo simulation (I)

- Introduce the notation
  \[ x_{0:k} = (x_0, \ldots, x_k) \]

- Represent posterior distribution using a set of samples or particles.
  \[
  \tilde{p}(x_{0:k} | D_k) = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_{0:k}^i} (dx_{0:k})
  \]

- Random samples \( x_{0:k}^i \) are drawn from the posterior distribution.
Perfect Monte Carlo simulation (II)

- Easy to approximate expectations of the form:

\[
E(g(x_{0:k})) = \int g(x_{0:k}) p(x_{0:k} \mid D_k) \, dx_{0:k}
\]

- by:

\[
\bar{E}(g(x_{0:k})) = \frac{1}{N} \sum_{i=1}^{N} g(x_{0:k}^i)
\]
Random samples and the pdf (I)

- Take $p(x) = \text{Gamma}(4,1)$
- Generate some random samples
- Plot histogram and basic approximation to pdf

200 samples

NCAF January Meeting, Aston University, Birmingham.
Random samples and the pdf (II)

500 samples

1000 samples

NCAF January Meeting, Aston University, Birmingham.
Random samples and the pdf (III)

5000 samples

200000 samples
Bayesian Importance Sampling (I)

Unfortunately it is often not possible to sample directly from the posterior distribution.

Circumvent by drawing from a known easy to sample proposal distribution \( q(x_{0:k} \mid D_k) \) giving:

\[
E(g(x_{0:k})) = \int g(x_{0:k}) \frac{p(x_{0:k} \mid D_k)}{q(x_{0:k} \mid D_k)} q(x_{0:k} \mid D_k) \, dx_{0:k}
\]

\[
= \int g(x_{0:k}) \frac{p(D_k \mid x_{0:k}) p(x_{0:k})}{p(D_k) q(x_{0:k} \mid D_k)} q(x_{0:k} \mid D_k) \, dx_{0:k}
\]

\[
= \int g(x_{0:k}) \frac{w_k(x_{0:k})}{p(D_k)} q(x_{0:k} \mid D_k) \, dx_{0:k}
\]
Bayesian Importance Sampling (II)

where $w_k(x_{0:k})$ are unnormalised importance weights:

$$w_k(x_{0:k}) = \frac{p(D_k \mid x_{0:k})p(x_{0:k})}{q(x_{0:k} \mid D_k)}$$

Now:

$$p(D_k) = \int p(D_k, x_{0:k}) \, dx_{0:k}$$

$$= \int \frac{p(D_k \mid x_{0:k})p(x_{0:k})q(x_{0:k} \mid D_k)}{q(x_{0:k} \mid D_k)} \, dx_{0:k}$$

$$= \int w_k(x_{0:k})q(x_{0:k} \mid D_k) \, dx_{0:k}$$

NCAF January Meeting, Aston University, Birmingham.
Bayesian Importance Sampling (III)

azzo Giving:

\[ E(g(x_{0:k})) = \frac{\int (g(x_{0:k})w_k(x_{0:k}))q(x_{0:k} \mid D_k) \, dx_{0:k}}{\int w_k(x_{0:k})q(x_{0:k} \mid D_k) \, dx_{0:k}} \]

azzo so that:

\[ E(g(x_{0:k})) = \frac{1}{N} \sum_{i=1}^{N} g(x_{0:k}^i)w_k(x_{0:k}^i) = \frac{1}{N} \sum_{i=1}^{N} w_k(x_{0:k}^i) \sum_{i=1}^{N} g(x_{0:k}^i)\tilde{w}_k(x_{0:k}^i) \]

azzo where \( \tilde{w}_k^i = \tilde{w}_k(x_{0:k}^i) \) are normalised importance weights

– and \( x_{0:k}^i \) are independent random samples from \( q(x_{0:k} \mid D_k) \)

NCAF January Meeting, Aston University, Birmingham.
Sequential Importance Sampling (I)

- Factorising the proposal distribution:

\[ q(x_{0:k} \mid D_k) = q(x_0) \prod_{j=1}^{k} q(x_j \mid x_{0:j-1}, D_j) \]

- and remembering that the state evolution is modelled as a Markov process

- obtain a recursive estimate of the importance weights:

\[ w_k = w_{k-1} \frac{p(y_k \mid x_k) p(x_k \mid x_{k-1})}{q(x_k \mid x_{0:k}, D_k)} \]
Derivation of SIR weights

Since:

\[ p(x_{0:k}) = p(x_0) \prod_{j=1}^{k} p(x_j \mid x_{j-1}) \quad \text{and} \quad p(D_k \mid x_{0:k}) = \prod_{j=1}^{k} p(y_j \mid x_j) \]

We have:

\[ w_k = \frac{p(D_k \mid x_{0:k}) p(x_{0:k})}{q(x_k \mid x_{0:k-1}, D_k) q(x_{0:k-1} \mid D_{k-1})} \]

\[ = w_{k-1} \frac{p(D_k \mid x_{0:k}) p(x_{0:k})}{p(D_{k-1} \mid x_{0:k-1}) p(x_{0:k-1}) q(x_k \mid x_{0:k-1}, D_k)} \]

\[ = w_{k-1} \frac{p(y_k \mid x_k) p(x_k \mid x_{k-1})}{q(x_k \mid x_{0:k-1}, D_k)} \]
Sequential Importance Sampling (II)

(choice of the proposal distribution):

\[ q(x_k | x_{0:k-1}, D_k) \]

Choose proposal function to minimize variance of \( w_k \) (Doucet et al. 1999):

\[ q(x_k | x_{0:k-1}, D_k) = p(x_k | x_{0:k-1}, D_k) \]

Although Common choice is the prior distribution:

\[ q(x_k | x_{0:k-1}, D_k) = p(x_k | x_{k-1}) \]
Sequential Importance Sampling (III)

Illustration of SIS:

\[ w_{t+1} = \frac{p(x_{t+1} | D_k)}{q(x_{t+1} | D_k)} w_t \]

Degeneracy problems:

- Variance of importance ratios \( p(x_{0:k} | D_k) / q(x_{0:k} | D_k) \) increases stochastically over time (Kong et al. 1994; Doucet et al. 1999).

NCAF January Meeting, Aston University, Birmingham.
SIS - why variance increase is bad

- Suppose we want to sample from the posterior
  - choose a proposal density to be very close to the posterior density
    - Then
      \[
      E_q \left( \frac{p(x_{0:k} \mid D_k)}{q(x_{0:k} \mid D_k)} \right) = 1
      \]
    - and
      \[
      \text{var}_q \left( \frac{p(x_{0:k} \mid D_k)}{q(x_{0:k} \mid D_k)} \right) = E_q \left( \left( \frac{p(x_{0:k} \mid D_k)}{q(x_{0:k} \mid D_k)} - 1 \right)^2 \right) = 0
      \]
  - So we expect the variance to be close to 0 to obtain reasonable estimates
    - thus a variance increase has a harmful effect on accuracy

NCAF January Meeting, Aston University, Birmingham.
Sequential Importance Sampling (IV)

Illustration of degeneracy:

NCAF January Meeting, Aston University, Birmingham.
Sampling-Importance Resampling

- SIS suffers from degeneracy problems so we don’t want to do that!
- Introduce a selection (resampling) step to eliminate samples with low importance ratios and multiply samples with high importance ratios.
- Resampling maps the weighted random measure $\{x_{0:k}^i, \tilde{w}_k(x_{0:k}^i)\}$ on to the equally weighted random measure $\{x_{0:k}^j, N^{-1}\}$
  - by sampling uniformly with replacement from $\{x_{0:k}^i; i = 1, \ldots, N\}$ with probabilities $\{\tilde{w}_k^i; i = 1, \ldots, N\}$
- Scheme generates $N_i$ children such that $\sum_{i=1}^{N} N_i = N$ and satisfies:
  - $E(N_i) = N\tilde{w}_k^i$
  - $\text{var}(N_i) = N\tilde{w}_k^i (1 - \tilde{w}_k^i)$
Improvements to SIR (I)

- Variety of resampling schemes with varying performance in terms of the variance of the particles \( \text{var}(N_i) \):
  - Systematic sampling (Carpenter et al., 1999).
  - Mixture of SIS and SIR, only resample when necessary (Liu & Chen, 1995; Doucet et al., 1999).

- Degeneracy may still be a problem:
  - During resampling a sample with high importance weight may be duplicated many times.
  - Samples may eventually collapse to a single point.
Improvements to SIR (II)

- To alleviate numerical degeneracy problems, sample smoothing methods may be adopted.
  
  - Roughening (Gordon et al., 1993).
    
    - Adds an independent jitter to the resampled particles
  
  - Prior boosting (Gordon et al., 1993).
    
    - Increase the number of samples from the proposal distribution to M>N,
    - but in the resampling stage only draw N particles.
Improvements to SIR (III)

- Local Monte Carlo methods for alleviating degeneracy:
  - Local linearisation - using an EKF (Doucet, 1999; Pitt & Shephard, 1999) or UKF (Doucet et al, 2000) to estimate the importance distribution.
  - Rejection methods (Müller, 1991; Doucet, 1999; Pitt & Shephard, 1999).
  - Auxiliary particle filters (Pitt & Shephard, 1999)
  - Kernel smoothing (Gordon, 1994; Hürzeler & Künsch, 1998; Liu & West, 2000; Musso et al., 2000).
  - MCMC methods (Müller, 1992; Gordon & Whitby, 1995; Berzuini et al., 1997; Gilks & Berzuini, 1998; Andrieu et al., 1999).
Improvements to SIR (IV)

- Illustration of SIR with sample smoothing:
MCMC move step

- Improve results by introducing MCMC steps with invariant distribution $p(x_{0:k} | D_k)$.
  - By applying a Markov transition kernel, the total variation of the current distribution w.r.t. the invariant distribution can only decrease.

- Introduces possibility of variable dimension state space through the use of reversible jump MCMC (de Freitas et al., 1999; Gilks & Berzuini, 2001)

NCAF January Meeting, Aston University, Birmingham.
Ingredients for SMC

- Importance sampling function
  - Gordon et al → \( p(x_k \mid x^i_{k-1}) \)
  - Optimal → \( p(x_k \mid x^{i}_{0:k-1}, D_k) \)
  - UKF → pdf from UKF at \( x^i_{k-1} \)

- Redistribution scheme
  - Gordon et al → SIR
  - Liu & Chen → Residual
  - Carpenter et al → Systematic
  - Liu & Chen, Doucet et al → Resample when necessary

- Careful initialisation procedure (for efficiency)

NCAF January Meeting, Aston University, Birmingham.
Basic Particle Filter - Schematic

Initialisation

\[ k = 0 \]

measurement

\[ y_k \]

Resampling

step

\{x_{0:k}^i, N^{-1}\}

Importance

sampling step

\{x_{0:k}^i, \tilde{w}_k(x_{0:k}^i)\}

Extract estimate, \( \hat{x}_{0:k} \)

NCAF January Meeting, Aston University, Birmingham.
Basic Particle Filter algorithm (I)

- Initialisation
  - \( k = 0 \)
  - For \( i = 1, \ldots, N \) sample \( x_0^i \sim p(x_0) \)
  - and set \( k = 1 \)

- In practice, to avoid having to take too many samples, for the first step we may want to ensure that we have a reasonable number of particles in the region of high likelihood
  - perhaps use MCMC techniques
Basic Particle Filter algorithm (II)

zu Importance Sampling step

- For $i = 1, \ldots, N$ sample $\tilde{x}_k^i \sim p(x_k \mid x_{k-1}^i)$
  and set $\tilde{x}_{0:k} = (x_{0:k-1}^i, x_k^i)$
- For $i = 1, \ldots, N$ evaluate the importance weights

  $w_k^i = p(y_k \mid \tilde{x}_k^i)$

- Normalise the importance weights,

  $\tilde{w}_k^i = w_k^i / \sum_{j=1}^{N} w_k^j$
Basic Particle Filter algorithm (III)

- **Resampling step**
  - Resample with replacement $N$ particles:
    $$(x^i_{0:k}; i = 1, \ldots, N)$$
  - from the set:
    $$(\tilde{x}^i_{0:k}; i = 1, \ldots, N)$$
  - according to the normalised importance weights, $\tilde{w}^i_k$

- **Set** $k \rightarrow k + 1$
  - proceed to the Importance Sampling step, as the next measurement arrives.
Example

- On-line Data Fusion (Marrs, 2000).
Example - Sensor Deployment

- Aim to reduce target sd below some threshold...
- … and keep it there
- … by placing the minimum number of sensors possible
- Sensor positions chosen according to particle distribution.

NCAF January Meeting, Aston University, Birmingham.
Example - In-situ monitoring of growing semiconductor crystal composition
Book Advert (or put this in or your fired)

  - Theoretical foundations - plus convergence proofs
  - Efficiency measures
  - Applications:
    - Target tracking; missile guidance; image tracking; terrain referenced navigation; exchange rate prediction; portfolio allocation; ellipsometry; electricity load forecasting; pollution monitoring; population biology; communications and audio engineering.
Conclusions

- On-line Bayesian learning a realistic proposition for many applications.
- Appropriate for complex non-linear/non-Gaussian models
  - don’t bother if KF based solution adequate.
- Representation of full posterior pdf leading to
  - estimation of moments.
  - estimation of HPD regions.
  - multi-modality easy to deal with.
- Model order can be included in unknowns.
- Can mix SMC and KF based solutions
Tracking Demo

- Illustrate a running particle filter
  - compare with Kalman Filter

- Running as we watch - not pre-recorded

- Pre-defined scenarios, or design your own
  - available to play with at coffee and lunch breaks.
2nd Book Advert

- Statistical Pattern Recognition
- Andrew Webb, DERA
- ISBN 0340741643,
- Paperback: 1999: £29.99
- Butterworth Heinemann

Contents:

- Introduction to SPR, Estimation, Density estimation, Linear discriminant analysis, Nonlinear discriminant analysis - neural networks, Nonlinear discriminant analysis - statistical methods, Classification trees, Feature selection and extraction, Clustering, Additional topics, Measures of dissimilarity, Parameter estimation, Linear algebra, Data, Probability theory.

NCAF January Meeting, Aston University, Birmingham.