Particle filters with independent resampling

Roland Lamberti[1], Yohan Petetin[1], François Septier[2] & François Desbouvries[1]

[1] Telecom Sudparis and CNRS UMR 5157, Evry, France

[2] Télécom Lille and CNRS UMR 9189, Villeneuve d'Ascq, France

41st ICASSP Conference, 20-25 March 2016, Shanghai, China

Outline

- Sequential Monte Carlo algorithms
- Distribution of the resampled particles
- Independent resampling
- Discussion
- Simulations

Sequential Monte Carlo algorithms

- Bayesian filtering
  $\{\mathbf{X}_k \in \mathbb{R}^p, \mathbf{Y}_k \in \mathbb{R}^q\}_{k\in\mathbb{N}}$ Hidden Markov Chain

$$p(\mathbf{x}_k|\mathbf{y}_{0:k}) = \frac{g_k(\mathbf{y}_k|\mathbf{x}_k)\int f_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{y}_{0:k-1})d\mathbf{x}_{k-1}}{p(\mathbf{y}_k|\mathbf{y}_{0:k-1})}$$

- In practice : Sequential Monte Carlo

  Propagate a set $\{\mathbf{x}^i_{0:k},\, w^i_k\}_{i=1}^N$ of weighted samples via sequential importance sampling

  $p(\mathbf{x}_k|\mathbf{y}_{0:k}) \leftarrow$ discrete approximation $\hat{p}(\mathbf{x}_k|\mathbf{y}_{0:k}) = \sum_{j=1}^N w^j_k \delta_{\mathbf{x}^j_k}$
  $\Theta_k = \int f(\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{0:k})\mathrm{d}\mathbf{x}_k \leftarrow \widehat{\Theta}_k = \sum_{j=1}^N w^j_k f(\mathbf{x}^j_k)$

The generic particle filter

- Sequential importance sampling

Sampling : sample $\tilde{\mathbf{x}}_k^i \sim q(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{y}_{0:k})$

Weighting : set $w_k^i \propto w_{k-1}^i \frac{f_{k|k-1}(\tilde{\mathbf{x}}_k^i|\mathbf{x}_{k-1}^i)g_k(\mathbf{y}_k|\tilde{\mathbf{x}}_k^i)}{q(\tilde{\mathbf{x}}_k^i|\mathbf{x}_{k-1}^i,\mathbf{y}_{0:k})}$, $\sum_{i=1}^N w_k^i = 1$,

$$\widehat{\Theta}_k^{SIS} = \sum_{i=1}^N w_k^i f(\tilde{\mathbf{x}}_k^i)$$

Resampling : sample $\mathbf{x}_k^i \sim \sum_{j=1}^N w_k^j \delta_{\tilde{\mathbf{x}}_k^j}$, set $w_k^i = 1/N$

$$\widehat{\Theta}_k^{SIR} = \sum_{i=1}^N \frac{1}{N} f(\mathbf{x}_k^i)$$

- The (optional) multinomial resampling step
  - fights against weight degeneracy
  - no local benefits : var$(\widehat{\Theta}_k^{SIR}) \geq$ var$(\widehat{\Theta}_k^{SIS})$

    but impacts subsequent performances (Cappé *et al.* 2005)
  - many variants (residual, stratified...)
    (Douc *et al.* 2005, Li *et al.* 2015)

The generic particle filter

- Sequential importance sampling

Sampling : sample $\tilde{\mathbf{x}}_k^i \sim q(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{y}_{0:k})$

Weighting : set $w_k^i \propto w_{k-1}^i \frac{f_{k|k-1}(\tilde{\mathbf{x}}_k^i|\mathbf{x}_{k-1}^i) g_k(\mathbf{y}_k|\tilde{\mathbf{x}}_k^i)}{q(\tilde{\mathbf{x}}_k^i|\mathbf{x}_{k-1}^i, \mathbf{y}_{0:k})}$, $\sum_{i=1}^N w_k^i = 1$,

$$\widehat{\Theta}_k^{SIS} = \sum_{i=1}^N w_k^i f(\tilde{\mathbf{x}}_k^i)$$

Resampling : sample $\mathbf{x}_k^i \sim \sum_{j=1}^N w_k^j \delta_{\tilde{\mathbf{x}}_k^j}$, set $w_k^i = 1/N$

$$\widehat{\Theta}_k^{SIR} = \sum_{i=1}^N \frac{1}{N} f(\mathbf{x}_k^i)$$

- The (optional) multinomial resampling step
  - fights against weight degeneracy
  - no local benefits : var$(\widehat{\Theta}_k^{SIR}) \geq$ var$(\widehat{\Theta}_k^{SIS})$

    but impacts subsequent performances (Cappé *et al.* 2005)
  - many variants (residual, stratified...)
    (Douc *et al.* 2005, Li *et al.* 2015)

- Sequential importance sampling

Sampling : sample $\tilde{\mathbf{x}}_k^i \sim q(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{y}_{0:k})$

Weighting : set $w_k^i \propto w_{k-1}^i \frac{f_{k|k-1}(\tilde{\mathbf{x}}_k^i|\mathbf{x}_{k-1}^i) g_k(\mathbf{y}_k|\tilde{\mathbf{x}}_k^i)}{q(\tilde{\mathbf{x}}_k^i|\mathbf{x}_{k-1}^i, \mathbf{y}_{0:k})}$, $\sum_{i=1}^N w_k^i = 1$,

$$\widehat{\Theta}_k^{SIS} = \sum_{i=1}^N w_k^i f(\tilde{\mathbf{x}}_k^i)$$

Resampling : sample $\mathbf{x}_k^i \sim \sum_{j=1}^N w_k^j \delta_{\tilde{\mathbf{x}}_k^j}$, set $w_k^i = 1/N$

$$\widehat{\Theta}_k^{SIR} = \sum_{i=1}^N \frac{1}{N} f(\mathbf{x}_k^i)$$

- The (optional) multinomial resampling step
  - fights against weight degeneracy
  - no local benefits : var$(\widehat{\Theta}_k^{SIR}) \geq$ var$(\widehat{\Theta}_k^{SIS})$

    but impacts subsequent performances (Cappé *et al.* 2005)
  - many variants (residual, stratified...)
    (Douc *et al.* 2005, Li *et al.* 2015)

The generic particle filter

- Sequential importance sampling

Sampling : sample $\tilde{\mathbf{x}}_k^i \sim q(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{y}_{0:k})$

Weighting : set $w_k^i \propto w_{k-1}^i \frac{f_{k|k-1}(\tilde{\mathbf{x}}_k^i|\mathbf{x}_{k-1}^i) g_k(\mathbf{y}_k|\tilde{\mathbf{x}}_k^i)}{q(\tilde{\mathbf{x}}_k^i|\mathbf{x}_{k-1}^i, \mathbf{y}_{0:k})}$, $\sum_{i=1}^N w_k^i = 1$,

$$\widehat{\Theta}_k^{SIS} = \sum_{i=1}^N w_k^i f(\tilde{\mathbf{x}}_k^i)$$

Resampling : sample $\mathbf{x}_k^i \sim \sum_{j=1}^N w_k^j \delta_{\tilde{\mathbf{x}}_k^j}$, set $w_k^i = 1/N$

$$\widehat{\Theta}_k^{SIR} = \sum_{i=1}^N \frac{1}{N} f(\mathbf{x}_k^i)$$

- The (optional) multinomial resampling step
  - fights against weight degeneracy
  - no local benefits : var$(\widehat{\Theta}_k^{SIR}) \geq$ var$(\widehat{\Theta}_k^{SIS})$

    but impacts subsequent performances (Cappé *et al.* 2005)
  - many variants (residual, stratified...)
    (Douc *et al.* 2005, Li *et al.* 2015)

Two observations

- Sequential importance sampling

Sampling : sample $\tilde{\mathbf{x}}_k^i \sim q(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{y}_{0:k})$

Weighting : set $w_k^i \propto w_{k-1}^i \frac{f_{k|k-1}(\tilde{\mathbf{x}}_k^i|\mathbf{x}_{k-1}^i) g_k(\mathbf{y}_k|\tilde{\mathbf{x}}_k^i)}{q(\tilde{\mathbf{x}}_k^i|\mathbf{x}_{k-1}^i, \mathbf{y}_{0:k})}$, $\sum_{i=1}^N w_k^i = 1$,

Resampling : sample $\mathbf{x}_k^i \sim \sum_{j=1}^N w_k^j \delta_{\tilde{\mathbf{x}}_k^j}$, set $w_k^i = 1/N$

-
  - Given $\{\mathbf{x}_{k-1}^i, w_{k-1}^i\}_{i=1}^N$, each $\mathbf{x}_k^i$ is drawn from

$$\tilde{q}(\mathbf{x}) = \sum_{i=1}^N \int \frac{p_i(\mathbf{x})}{\frac{p_i(\mathbf{x})}{q_i(\mathbf{x})} + \sum_{j \neq i} \frac{p_j(\mathbf{x}^j)}{q_j(\mathbf{x}^j)}} \prod_{j \neq i} q_j(\mathbf{x}^j) \mathrm{d}\mathbf{x}^1 \cdots \mathbf{x}^j$$

where $p_i(\mathbf{x}) = w_{k-1}^i f_{k|k-1}(\mathbf{x}|\mathbf{x}_{k-1}^i) g_k(\mathbf{y}_k|\mathbf{x})$,
$q_i(\mathbf{x}) = q(\mathbf{x}|\mathbf{x}_{k-1}^i, \mathbf{y}_{0:k})$;

  - $\{\mathbf{x}_k^i\}_{i=1}^N$ are independent given $\{\mathbf{x}_{k-1}^i, w_{k-1}^i\}_{i=1}^N$ *and* $\{\tilde{\mathbf{x}}_k^i\}_{i=1}^N$;
    $\{\mathbf{x}_k^i\}_{i=1}^N$ are **dependent** given $\{\mathbf{x}_{k-1}^i, w_{k-1}^i\}_{i=1}^N$ only.

Two observations

- Sequential importance sampling

  Sampling : sample $\tilde{\mathbf{x}}_k^i \sim q(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{y}_{0:k})$

  Weighting : set $w_k^i \propto w_{k-1}^i \frac{f_{k|k-1}(\tilde{\mathbf{x}}_k^i|\mathbf{x}_{k-1}^i) g_k(\mathbf{y}_k|\tilde{\mathbf{x}}_k^i)}{q(\tilde{\mathbf{x}}_k^i|\mathbf{x}_{k-1}^i, \mathbf{y}_{0:k})}$, $\sum_{i=1}^N w_k^i = 1$,

  Resampling : sample $\mathbf{x}_k^i \sim \sum_{j=1}^N w_k^j \delta_{\tilde{\mathbf{x}}_k^j}$, set $w_k^i = 1/N$

- • Given $\{\mathbf{x}_{k-1}^i, w_{k-1}^i\}_{i=1}^N$, each $\mathbf{x}_k^i$ is drawn from

$$\tilde{q}(\mathbf{x}) = \sum_{i=1}^N \int \frac{p_i(\mathbf{x})}{\frac{p_i(\mathbf{x})}{q_i(\mathbf{x})} + \sum_{j \neq i} \frac{p_j(\mathbf{x}^j)}{q_j(\mathbf{x}^j)}} \prod_{j \neq i} q_j(\mathbf{x}^j) \mathrm{d}\mathbf{x}^1 \cdots \mathbf{x}^j$$

where
$$p_i(\mathbf{x}) = w_{k-1}^i f_{k|k-1}(\mathbf{x}|\mathbf{x}_{k-1}^i) g_k(\mathbf{y}_k|\mathbf{x}),$$
$$q_i(\mathbf{x}) = q(\mathbf{x}|\mathbf{x}_{k-1}^i, \mathbf{y}_{0:k});$$

• $\{\mathbf{x}_k^i\}_{i=1}^N$ are independent given $\{\mathbf{x}_{k-1}^i, w_{k-1}^i\}_{i=1}^N$ *and* $\{\tilde{\mathbf{x}}_k^i\}_{i=1}^N$;
$\{\mathbf{x}_k^i\}_{i=1}^N$ are **dependent** given $\{\mathbf{x}_{k-1}^i, w_{k-1}^i\}_{i=1}^N$ only.

- Sequential importance sampling

Sampling : sample $\tilde{\mathbf{x}}_k^i \sim q(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{y}_{0:k})$

Weighting : set $w_k^i \propto w_{k-1}^i \frac{f_{k|k-1}(\tilde{\mathbf{x}}_k^i|\mathbf{x}_{k-1}^i) g_k(\mathbf{y}_k|\tilde{\mathbf{x}}_k^i)}{q(\tilde{\mathbf{x}}_k^i|\mathbf{x}_{k-1}^i, \mathbf{y}_{0:k})}$, $\sum_{i=1}^N w_k^i = 1$,

Resampling : sample $\mathbf{x}_k^i \sim \sum_{j=1}^N w_k^j \delta_{\tilde{\mathbf{x}}_k^j}$, set $w_k^i = 1/N$

-
  - Given $\{\mathbf{x}_{k-1}^i, w_{k-1}^i\}_{i=1}^N$, each $\mathbf{x}_k^i$ is drawn from

$$\tilde{q}(\mathbf{x}) = \sum_{i=1}^N \int \frac{p_i(\mathbf{x})}{\frac{p_i(\mathbf{x})}{q_i(\mathbf{x})} + \sum_{j \neq i} \frac{p_j(\mathbf{x}^j)}{q_j(\mathbf{x}^j)}} \prod_{j \neq i} q_j(\mathbf{x}^j) \mathrm{d}\mathbf{x}^1 \cdots \mathbf{x}^j$$

where
$$p_i(\mathbf{x}) = w_{k-1}^i f_{k|k-1}(\mathbf{x}|\mathbf{x}_{k-1}^i) g_k(\mathbf{y}_k|\mathbf{x}),$$
$$q_i(\mathbf{x}) = q(\mathbf{x}|\mathbf{x}_{k-1}^i, \mathbf{y}_{0:k});$$

  - $\{\mathbf{x}_k^i\}_{i=1}^N$ are independent given $\{\mathbf{x}_{k-1}^i, w_{k-1}^i\}_{i=1}^N$ *and* $\{\tilde{\mathbf{x}}_k^i\}_{i=1}^N$;
    $\{\mathbf{x}_k^i\}_{i=1}^N$ are **dependent** given $\{\mathbf{x}_{k-1}^i, w_{k-1}^i\}_{i=1}^N$ only.

Independent resampling

$$\tilde{q}(\mathbf{x}) = \sum_{i=1}^{N} \int \frac{p_i(\mathbf{x})}{\frac{p_i(\mathbf{x})}{q_i(\mathbf{x})} + \sum_{j \neq i} \frac{p_j(\mathbf{x}^j)}{q_j(\mathbf{x}^j)}} \prod_{j \neq i} q_j(\mathbf{x}^j) \mathrm{d}\mathbf{x}^1 \cdots \mathbf{x}^j$$

where $p_i(\mathbf{x}) = w^i_{k-1} f_{k|k-1}(\mathbf{x}|\mathbf{x}^i_{k-1}) g_k(\mathbf{y}_k|\mathbf{x})$,
$q_i(\mathbf{x}) = q(\mathbf{x}|\mathbf{x}^i_{k-1}, \mathbf{y}_{0:k})$;

- Modify resampling step, s.t. resampled particles are drawn i.i.d. from $\tilde{q}$?
- How to sample i.i.d. from $\tilde{q}$?
- Potential benefits ?

PF with ind. resampling - optimal CID

$$\tilde{q}(\mathbf{x}) = \sum_{i=1}^{N} \int \frac{p_i(\mathbf{x})}{\frac{p_i(\mathbf{x})}{q_i(\mathbf{x})} + \sum_{j\neq i} \frac{p_j(\mathbf{x}^j)}{q_j(\mathbf{x}^j)}} \prod_{j\neq i} q_j(\mathbf{x}^j)\mathrm{d}\mathbf{x}^1 \cdots \mathbf{x}^j$$

where $p_i(\mathbf{x}) = w_{k-1}^i f_{k|k-1}(\mathbf{x}|\mathbf{x}_{k-1}^i) g_k(\mathbf{y}_k|\mathbf{x})$,
$q_i(\mathbf{x}) = q(\mathbf{x}|\mathbf{x}_{k-1}^i, \mathbf{y}_{0:k})$;

- Optimal CID : $q_i(\mathbf{x}_k) = p(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{y}_k)$, $w_k^i \propto w_{k-1}^i p(\mathbf{y}_k|\mathbf{x}_{k-1}^i)$
$\tilde{q}$ reduces to mixture pdf $\tilde{q}(\mathbf{x}_k) = \sum_{i=1}^N w_k^i p(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{y}_k)$
- i.i.d. sampling from a mixture $\sum_{i=1}^N w_k^i p_i(x)$ is simple and fast
  - **Weighting** : set $w_k^i \propto w_{k-1}^i p(\mathbf{y}_k|\mathbf{x}_{k-1}^i)$, $\sum_{i=1}^N w_k^i = 1$
  - **Resampling** : sample $\tilde{\mathbf{x}}_{k-1}^i \sim \sum_{j=1}^N w_k^j \delta_{\mathbf{x}_{k-1}^j}$
  - **Sampling** : sample $\mathbf{x}_k^i \sim q(\mathbf{x}_k|\tilde{\mathbf{x}}_{k-1}^i, \mathbf{y}_{0:k})$
- Fully-adapted APF (Pitt & Shephard 1999) outperforms optimal SIR (Cappé *et al.* 2005, Petetin *et al.* 2013)

PF with ind. resampling - optimal CID

$$\tilde{q}(\mathbf{x}) = \sum_{i=1}^{N} \int \frac{p_i(\mathbf{x})}{\frac{p_i(\mathbf{x})}{q_i(\mathbf{x})} + \sum_{j\neq i} \frac{p_j(\mathbf{x}^j)}{q_j(\mathbf{x}^j)}} \prod_{j\neq i} q_j(\mathbf{x}^j) \mathrm{d}\mathbf{x}^1 \cdots \mathbf{x}^j$$

where $p_i(\mathbf{x}) = w_{k-1}^i f_{k|k-1}(\mathbf{x}|\mathbf{x}_{k-1}^i) g_k(\mathbf{y}_k|\mathbf{x})$,
$q_i(\mathbf{x}) = q(\mathbf{x}|\mathbf{x}_{k-1}^i, \mathbf{y}_{0:k})$;

- Optimal CID : $q_i(\mathbf{x}_k) = p(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{y}_k)$, $w_k^i \propto w_{k-1}^i p(\mathbf{y}_k|\mathbf{x}_{k-1}^i)$
  $\tilde{q}$ reduces to mixture pdf $\tilde{q}(\mathbf{x}_k) = \sum_{i=1}^N w_k^i p(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{y}_k)$
- i.i.d. sampling from a mixture $\sum_{i=1}^N w_k^i p_i(x)$ is simple and fast
  Weighting : set $w_k^i \propto w_{k-1}^i p(\mathbf{y}_k|\mathbf{x}_{k-1}^i)$, $\sum_{i=1}^N w_k^i = 1$
  Resampling : sample $\tilde{\mathbf{x}}_{k-1}^i \sim \sum_{j=1}^N w_k^j \delta_{\mathbf{x}_{k-1}^j}$
  Sampling : sample $\mathbf{x}_k^i \sim q(\mathbf{x}_k|\tilde{\mathbf{x}}_{k-1}^i, \mathbf{y}_{0:k})$
- Fully-adapted APF (Pitt & Shephard 1999) outperforms optimal SIR (Cappé *et al.* 2005, Petetin *et al.* 2013)

PF with ind. resampling - optimal CID

$$\tilde{q}(\mathbf{x}) = \sum_{i=1}^{N} \int \frac{p_i(\mathbf{x})}{\frac{p_i(\mathbf{x})}{q_i(\mathbf{x})} + \sum_{j\neq i} \frac{p_j(\mathbf{x}^j)}{q_j(\mathbf{x}^j)}} \prod_{j\neq i} q_j(\mathbf{x}^j) \mathrm{d}\mathbf{x}^1 \cdots \mathbf{x}^j$$

where $p_i(\mathbf{x}) = w_{k-1}^i f_{k|k-1}(\mathbf{x}|\mathbf{x}_{k-1}^i) g_k(\mathbf{y}_k|\mathbf{x})$,
$q_i(\mathbf{x}) = q(\mathbf{x}|\mathbf{x}_{k-1}^i, \mathbf{y}_{0:k})$;

- Optimal CID : $q_i(\mathbf{x}_k) = p(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{y}_k)$, $w_k^i \propto w_{k-1}^i p(\mathbf{y}_k|\mathbf{x}_{k-1}^i)$
  $\tilde{q}$ reduces to mixture pdf $\tilde{q}(\mathbf{x}_k) = \sum_{i=1}^N w_k^i p(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{y}_k)$
- i.i.d. sampling from a mixture $\sum_{i=1}^N w_k^i p_i(x)$ is simple and fast
  Weighting : set $w_k^i \propto w_{k-1}^i p(\mathbf{y}_k|\mathbf{x}_{k-1}^i)$, $\sum_{i=1}^N w_k^i = 1$
  Resampling : sample $\tilde{\mathbf{x}}_{k-1}^i \sim \sum_{j=1}^N w_k^j \delta_{\mathbf{x}_{k-1}^j}$
  Sampling : sample $\mathbf{x}_k^i \sim q(\mathbf{x}_k|\tilde{\mathbf{x}}_{k-1}^i, \mathbf{y}_{0:k})$
- Fully-adapted APF (Pitt & Shephard 1999) outperforms optimal SIR (Cappé *et al.* 2005, Petetin *et al.* 2013)

PF with ind. resampling - optimal CID

$$\tilde{q}(\mathbf{x}) = \sum_{i=1}^{N} \int \frac{p_i(\mathbf{x})}{\frac{p_i(\mathbf{x})}{q_i(\mathbf{x})} + \sum_{j \neq i} \frac{p_j(\mathbf{x}^j)}{q_j(\mathbf{x}^j)}} \prod_{j \neq i} q_j(\mathbf{x}^j) \mathrm{d}\mathbf{x}^1 \cdots \mathbf{x}^j$$

where $p_i(\mathbf{x}) = w_{k-1}^i f_{k|k-1}(\mathbf{x}|\mathbf{x}_{k-1}^i) g_k(\mathbf{y}_k|\mathbf{x})$,
$q_i(\mathbf{x}) = q(\mathbf{x}|\mathbf{x}_{k-1}^i, \mathbf{y}_{0:k})$;

- Optimal CID : $q_i(\mathbf{x}_k) = p(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{y}_k)$, $w_k^i \propto w_{k-1}^i p(\mathbf{y}_k|\mathbf{x}_{k-1}^i)$
  $\tilde{q}$ reduces to mixture pdf $\tilde{q}(\mathbf{x}_k) = \sum_{i=1}^N w_k^i p(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{y}_k)$
- i.i.d. sampling from a mixture $\sum_{i=1}^N w_k^i p_i(x)$ is simple and fast
  Weighting : set $w_k^i \propto w_{k-1}^i p(\mathbf{y}_k|\mathbf{x}_{k-1}^i)$, $\sum_{i=1}^N w_k^i = 1$
  Resampling : sample $\tilde{\mathbf{x}}_{k-1}^i \sim \sum_{j=1}^N w_k^j \delta_{\mathbf{x}_{k-1}^j}$
  Sampling : sample $\mathbf{x}_k^i \sim q(\mathbf{x}_k|\tilde{\mathbf{x}}_{k-1}^i, \mathbf{y}_{0:k})$
- Fully-adapted APF (Pitt & Shephard 1999) outperforms optimal SIR (Cappé *et al.* 2005, Petetin *et al.* 2013)

PF with ind. resampling - general case

$$\tilde{q}(\mathbf{x}) = \sum_{i=1}^{N} \int \frac{p_i(\mathbf{x})}{\frac{p_i(\mathbf{x})}{q_i(\mathbf{x})} + \sum_{j \neq i} \frac{p_j(\mathbf{x}^j)}{q_j(\mathbf{x}^j)}} \prod_{j \neq i} q_j(\mathbf{x}^j) \mathrm{d}\mathbf{x}^1 \cdots \mathbf{x}^j$$

where $p_i(\mathbf{x}) = w_{k-1}^i f_{k|k-1}(\mathbf{x}|\mathbf{x}_{k-1}^i) g_k(\mathbf{y}_k|\mathbf{x}),$
$q_i(\mathbf{x}) = q(\mathbf{x}|\mathbf{x}_{k-1}^i, \mathbf{y}_{0:k});$

- $p(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{y}_k)$ nor $p(\mathbf{y}_k|\mathbf{x}_{k-1}^i)$ computable in most models
- $\tilde{q}$ is still a mixture, but components cannot be computed
- PF with **independent resampling** : for all $1 \leq i, j \leq M$
    - Sampling : sample $\tilde{\mathbf{x}}_k^{i,j} \sim q(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{y}_{0:k})$
    - Weighting : set $w_k^{i,j} \propto w_{k-1}^i \times \frac{f_{k|k-1}(\tilde{\mathbf{x}}_k^{i,j}|\mathbf{x}_{k-1}^i) g_k(\mathbf{y}_k|\tilde{\mathbf{x}}_k^{i,j})}{q(\tilde{\mathbf{x}}_k^{i,j}|\mathbf{x}_{k-1}^i, \mathbf{y}_{0:k})}$, $\sum_{i=1}^{M} w_k^{i,j} = 1$
    - Resampling : sample $\mathbf{x}_k^j \sim \sum_{l=1}^{M} w_k^{l,j} \delta_{\tilde{\mathbf{x}}_k^{l,j}}$, set $w_k^j = 1/M$

PF with ind. resampling - general case

$$\tilde{q}(\mathbf{x}) = \sum_{i=1}^{N} \int \frac{p_i(\mathbf{x})}{\frac{p_i(\mathbf{x})}{q_i(\mathbf{x})} + \sum_{j \neq i} \frac{p_j(\mathbf{x}^j)}{q_j(\mathbf{x}^j)}} \prod_{j \neq i} q_j(\mathbf{x}^j) \mathrm{d}\mathbf{x}^1 \cdots \mathbf{x}^j$$

where $p_i(\mathbf{x}) = w_{k-1}^i f_{k|k-1}(\mathbf{x}|\mathbf{x}_{k-1}^i) g_k(\mathbf{y}_k|\mathbf{x})$,
$q_i(\mathbf{x}) = q(\mathbf{x}|\mathbf{x}_{k-1}^i, \mathbf{y}_{0:k})$;

- $p(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{y}_k)$ nor $p(\mathbf{y}_k|\mathbf{x}_{k-1}^i)$ computable in most models
- $\tilde{q}$ is still a mixture, but components cannot be computed
- PF with **independent resampling** : for all $1 \leq i, j \leq M$
  - Sampling : sample $\tilde{\mathbf{x}}_k^{i,j} \sim q(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{y}_{0:k})$
  - Weighting : set $w_k^{i,j} \propto w_{k-1}^i \times \frac{f_{k|k-1}(\tilde{\mathbf{x}}_k^{i,j}|\mathbf{x}_{k-1}^i) g_k(\mathbf{y}_k|\tilde{\mathbf{x}}_k^{i,j})}{q(\tilde{\mathbf{x}}_k^{i,j}|\mathbf{x}_{k-1}^i, \mathbf{y}_{0:k})}$, $\sum_{i=1}^{M} w_k^{i,j} = 1$
  - Resampling : sample $\mathbf{x}_k^j \sim \sum_{l=1}^{M} w_k^{l,j} \delta_{\tilde{\mathbf{x}}_k^{l,j}}$, set $w_k^j = 1/M$

PF with ind. resampling - general case

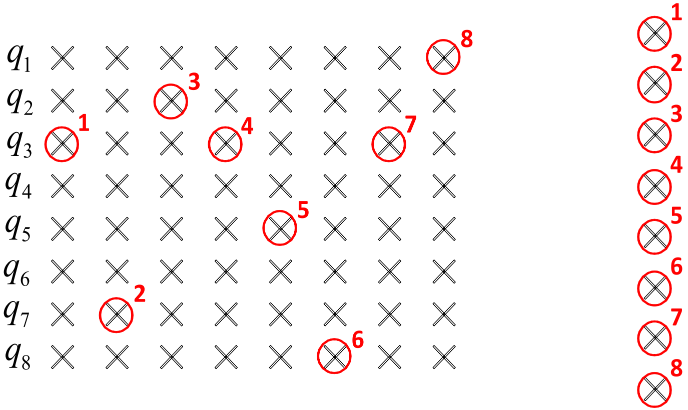
$$\tilde{q}(\mathbf{x}) = \sum_{i=1}^{N} \int \frac{p_i(\mathbf{x})}{\frac{p_i(\mathbf{x})}{q_i(\mathbf{x})} + \sum_{j \neq i} \frac{p_j(\mathbf{x}^j)}{q_j(\mathbf{x}^j)}} \prod_{j \neq i} q_j(\mathbf{x}^j) \mathrm{d}\mathbf{x}^1 \cdots \mathbf{x}^j$$

where $p_i(\mathbf{x}) = w_{k-1}^i f_{k|k-1}(\mathbf{x}|\mathbf{x}_{k-1}^i) g_k(\mathbf{y}_k|\mathbf{x})$,
$q_i(\mathbf{x}) = q(\mathbf{x}|\mathbf{x}_{k-1}^i, \mathbf{y}_{0:k})$;

- $p(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{y}_k)$ nor $p(\mathbf{y}_k|\mathbf{x}_{k-1}^i)$ computable in most models
- $\tilde{q}$ is still a mixture, but components cannot be computed
- PF with **independent resampling** : for all $1 \leq i, j \leq M$
  - Sampling : sample $\tilde{\mathbf{x}}_k^{i,j} \sim q(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{y}_{0:k})$
  - Weighting : set $w_k^{i,j} \propto w_{k-1}^i \times \frac{f_{k|k-1}(\tilde{\mathbf{x}}_k^{i,j}|\mathbf{x}_{k-1}^i) g_k(\mathbf{y}_k|\tilde{\mathbf{x}}_k^{i,j})}{q(\tilde{\mathbf{x}}_k^{i,j}|\mathbf{x}_{k-1}^i, \mathbf{y}_{0:k})}$, $\sum_{i=1}^{M} w_k^{i,j} = 1$
  - Resampling : sample $\mathbf{x}_k^j \sim \sum_{l=1}^{M} w_k^{l,j} \delta_{\tilde{\mathbf{x}}_k^{l,j}}$, set $w_k^j = 1/M$

PF with independent resampling

Example with $M = 8$: we obtain 8 particles drawn i.i.d. according to $\tilde{q}_8$



Independent resampling : discussion

- Post-resampling PF estimators with dependent samples $\widehat{\Theta}_k$ (with independent samples $\widetilde{\Theta}_k$) : $\widehat{\Theta}_k / \widetilde{\Theta}_k = \frac{1}{N} \sum_{i=1}^{N} f(\mathbf{x}_k^i)$

$$\begin{aligned} \mathrm{E}(\widehat{\Theta}_k) &= \mathrm{E}(\widetilde{\Theta}_k), \\ \mathrm{var}(\widehat{\Theta}_k) &= \mathrm{var}(\widetilde{\Theta}_k) + \frac{N-1}{N} \mathrm{var}\left(\sum_{i=1}^{N} w_k^i f(\tilde{\mathbf{x}}_k^i) \right) \end{aligned}$$

- No support degeneracy : better particle diversity for the next iteration

- M^2 Sampling and weighting steps : higher computational cost than the classical PF if $M = N$.
  However
  - Resampling is not necessarily needed at each iteration
  - Independent resampling can be parallelized
  - In some cases, performs better even when $M^2 + M = 2N$

Independent resampling : discussion

- Post-resampling PF estimators with dependent samples $\widehat{\Theta}_k$ (with independent samples $\widetilde{\Theta}_k$) : $\widehat{\Theta}_k/\widetilde{\Theta}_k = \frac{1}{N}\sum_{i=1}^{N} f(\mathbf{x}_k^i)$

$$\begin{aligned}
\mathrm{E}(\widehat{\Theta}_k) &= \mathrm{E}(\widetilde{\Theta}_k), \\
\mathrm{var}(\widehat{\Theta}_k) &= \mathrm{var}(\widetilde{\Theta}_k) + \frac{N-1}{N}\mathrm{var}\left(\sum_{i=1}^{N} w_k^i f(\tilde{\mathbf{x}}_k^i)\right)
\end{aligned}$$

- No support degeneracy : better particle diversity for the next iteration

- M^2 Sampling and weighting steps : higher computational cost than the classical PF if $M = N$.
  However
  - Resampling is not necessarily needed at each iteration
  - Independent resampling can be parallelized
  - In some cases, performs better even when $M^2 + M = 2N$

Independent resampling : discussion

- Post-resampling PF estimators with dependent samples $\widehat{\Theta}_k$ (with independent samples $\widetilde{\Theta}_k$) : $\widehat{\Theta}_k/\widetilde{\Theta}_k = \frac{1}{N}\sum_{i=1}^{N} f(\mathbf{x}_k^i)$

$$\begin{aligned} \mathrm{E}(\widehat{\Theta}_k) &= \mathrm{E}(\widetilde{\Theta}_k), \\ \mathrm{var}(\widehat{\Theta}_k) &= \mathrm{var}(\widetilde{\Theta}_k) + \frac{N-1}{N}\mathrm{var}\left(\sum_{i=1}^{N} w_k^i f(\tilde{\mathbf{x}}_k^i)\right) \end{aligned}$$

- No support degeneracy : better particle diversity for the next iteration

- M^2 Sampling and weighting steps : higher computational cost than the classical PF if $M = N$.
  However
  - Resampling is not necessarily needed at each iteration
  - Independent resampling can be parallelized
  - In some cases, performs better even when $M^2 + M = 2N$

Independent resampling : discussion

- Post-resampling PF estimators with dependent samples $\widehat{\Theta}_k$ (with independent samples $\widetilde{\Theta}_k$) : $\widehat{\Theta}_k/\widetilde{\Theta}_k = \frac{1}{N}\sum_{i=1}^{N} f(\mathbf{x}_k^i)$

$$\begin{aligned} \mathrm{E}(\widehat{\Theta}_k) &= \mathrm{E}(\widetilde{\Theta}_k), \\ \mathrm{var}(\widehat{\Theta}_k) &= \mathrm{var}(\widetilde{\Theta}_k) + \frac{N-1}{N}\mathrm{var}\left(\sum_{i=1}^{N} w_k^i f(\tilde{\mathbf{x}}_k^i)\right) \end{aligned}$$

- No support degeneracy : better particle diversity for the next iteration

- M^2 Sampling and weighting steps : higher computational cost than the classical PF if $M = N$.
  However
  - Resampling is not necessarily needed at each iteration
  - Independent resampling can be parallelized
  - In some cases, performs better even when $M^2 + M = 2N$

- Post-resampling PF estimators with dependent samples $\widehat{\Theta}_k$ (with independent samples $\widetilde{\Theta}_k$) : $\widehat{\Theta}_k / \widetilde{\Theta}_k = \frac{1}{N}\sum_{i=1}^{N} f(\mathbf{x}_k^i)$

$$\begin{aligned} \mathrm{E}(\widehat{\Theta}_k) &= \mathrm{E}(\widetilde{\Theta}_k), \\ \mathrm{var}(\widehat{\Theta}_k) &= \mathrm{var}(\widetilde{\Theta}_k) + \frac{N-1}{N}\mathrm{var}\left(\sum_{i=1}^{N} w_k^i f(\tilde{\mathbf{x}}_k^i)\right) \end{aligned}$$

- No support degeneracy : better particle diversity for the next iteration

- M^2 Sampling and weighting steps : higher computational cost than the classical PF if $M = N$.
  However
  - Resampling is not necessarily needed at each iteration
  - Independent resampling can be parallelized
  - In some cases, performs better even when $M^2 + M = 2N$

Independent resampling : discussion

- Post-resampling PF estimators with dependent samples $\widehat{\Theta}_k$ (with independent samples $\widetilde{\Theta}_k$) : $\widehat{\Theta}_k / \widetilde{\Theta}_k = \frac{1}{N} \sum_{i=1}^{N} f(\mathbf{x}_k^i)$

$$
\begin{aligned}
\mathrm{E}(\widehat{\Theta}_k) &= \mathrm{E}(\widetilde{\Theta}_k), \\
\mathrm{var}(\widehat{\Theta}_k) &= \mathrm{var}(\widetilde{\Theta}_k) + \frac{N-1}{N} \mathrm{var}\left(\sum_{i=1}^{N} w_k^i f(\tilde{\mathbf{x}}_k^i) \right)
\end{aligned}
$$

- No support degeneracy : better particle diversity for the next iteration

- M^2 Sampling and weighting steps : higher computational cost than the classical PF if $M = N$.
  However
  - Resampling is not necessarily needed at each iteration
  - Independent resampling can be parallelized
  - In some cases, performs better even when $M^2 + M = 2N$

Connection with existing works

- **Island particle filtering** (Vergé *et al.* 2015) :
  - Divides a set of N particles into N_1 islands of N_2 particles each
  - Resampling at the island level can reduce the bias introduced by this division
  - Parallelizable
  - Resampling still produces dependent draws

- **The Nested SMC algorithm** (Naesseth *et al.* 2015, Jaoua *et al.* 2013) :
  - Empirical approximation of optimal conditional importance distribution and predictive likelihood, then use of an FA-APF algorithm
  - Can generate duplicate particles since $\hat{q}^{\mathrm{opt}}$ is discrete

Simulations : Polar target tracking model

- Non-linear target tracking with range-bearing measurements
- Cartesian coordinates $\mathbf{x}_k = [p_{x,k}, \dot{p}_{x,k}, p_{y,k}, \dot{p}_{y,k}]^T$

$$\begin{aligned}
\mathbf{x}_k &= \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{u}_k \\
\mathbf{y}_k &= \begin{pmatrix} \sqrt{p_{x,k}^2 + p_{y,k}^2} \\ \arctan \frac{p_{y,k}}{p_{x,k}} \end{pmatrix} + \mathbf{v}_k
\end{aligned}$$

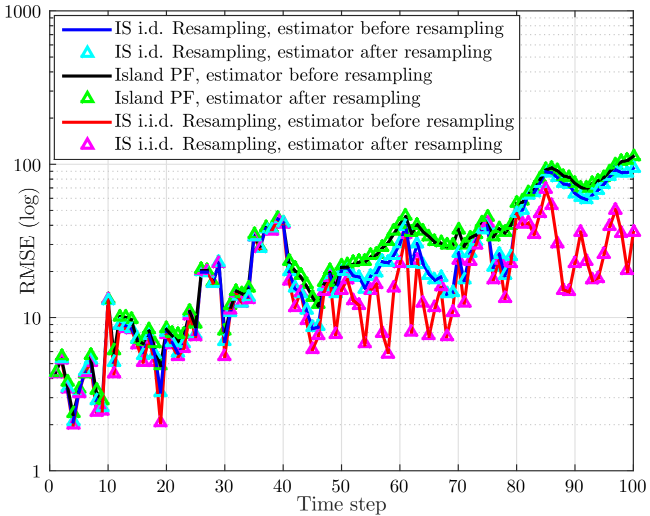
$\mathbf{x}_0, \mathbf{u}_1, \cdots, \mathbf{u}_k, \mathbf{v}_0, \cdots, \mathbf{y}_k$ ind., $\mathbf{u}_k \sim \mathcal{N}(\mathbf{0}_4, \mathbf{Q}), \mathbf{v}_k \sim \mathcal{N}(\mathbf{0}_2, \mathbf{R})$,

$$\mathbf{Q} = \sigma_Q^2 \begin{pmatrix} \frac{\tau^3}{3} & \frac{\tau^2}{2} & 0 & 0 \\ \frac{\tau^2}{2} & \tau & 0 & 0 \\ 0 & 0 & \frac{\tau^3}{3} & \frac{\tau^2}{2} \\ 0 & 0 & \frac{\tau^2}{2} & \tau \end{pmatrix}, \mathbf{R} = \begin{pmatrix} \sigma_\rho^2 & 0 \\ 0 & \sigma_\theta^2 \end{pmatrix}, \tau = 1.$$

- RMSE of the estimators averaged over $N_{\mathrm{MC}} = 100$ MC runs

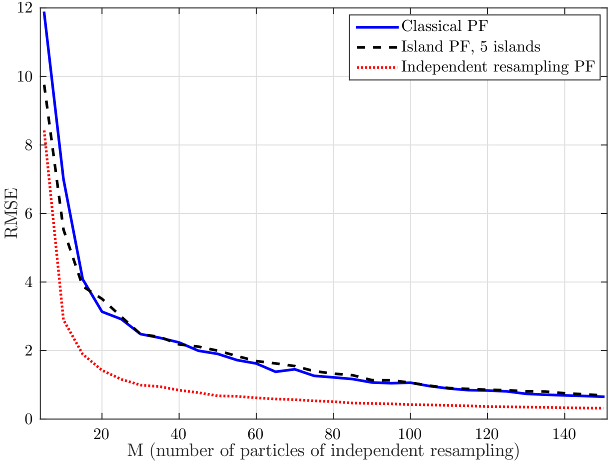
Independent resampling : $M = N$

$M = N = 500$ particles ; $\sigma_Q = \sqrt{10}, \sigma_\rho = 1$ and $\sigma_\theta = \frac{\pi}{180}$; 5 islands



Independent resampling : $M^2 + M = 2N$

$\sigma_Q = \sqrt{10}$, $\sigma_\rho = 0.05$ and $\sigma_\theta = \frac{\pi}{3600}$
PF : $N = (M^2 + M)/2$; IPF : $5 \times [(M^2 + M)/10]$; Ind. PF : M.



Conclusions

- We propose an independent resampling scheme for particle filtering that produces conditionally independent draws from the same distribution as that induced by multinomial resampling
- The algorithm is parallelizable and ensures better particle diversity
- It yields better performance than a classical (dependent) PF at even lower sampling cost in informative measurement scenarios

Conclusions

- We propose an independent resampling scheme for particle filtering that produces conditionally independent draws from the same distribution as that induced by multinomial resampling
- The algorithm is parallelizable and ensures better particle diversity
- It yields better performance than a classical (dependent) PF at even lower sampling cost in informative measurement scenarios

Conclusions

- We propose an independent resampling scheme for particle filtering that produces conditionally independent draws from the same distribution as that induced by multinomial resampling
- The algorithm is parallelizable and ensures better particle diversity
- It yields better performance than a classical (dependent) PF at even lower sampling cost in informative measurement scenarios