Revisit of the resampling mechanism used in importance sampling methods

François Septier
work with R. Lamberti, Y. Petetin and F. Desbouvries

STM2018 at ISM
28th February 2018

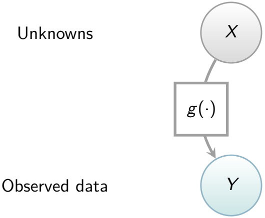
IMT Lille Douai
École Mines-Télécom
IMT-Université de Lille

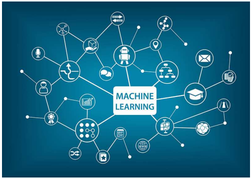
CRIStAL
Centre de Recherche en Informatique,
Signal et Automatique de Lille

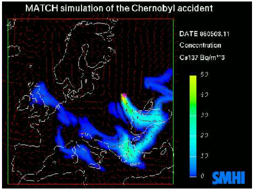
Outline

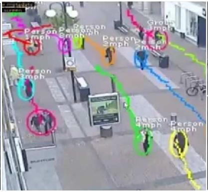
General context

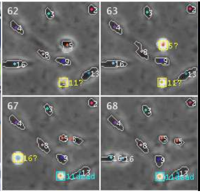
In many applications $\rightsquigarrow$ interest is in learning about unknowns from observed (noisy) data

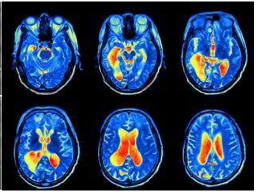




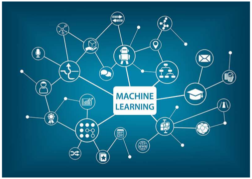




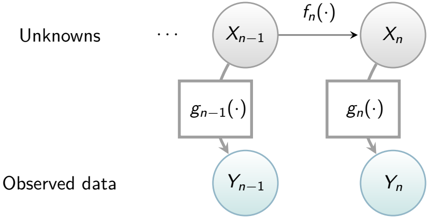


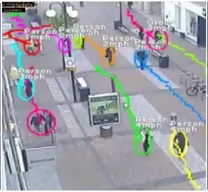


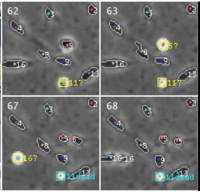
General context

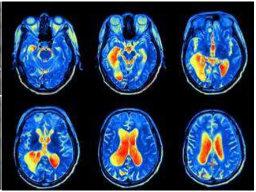


In many applications $\rightsquigarrow$ interest is in learning about unknowns from observed (noisy) data



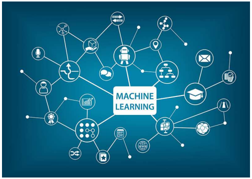


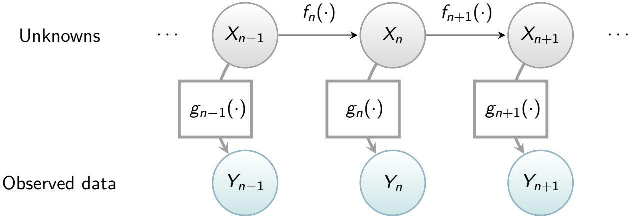


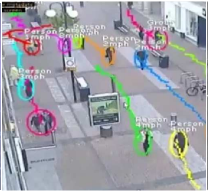


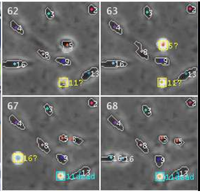
General context

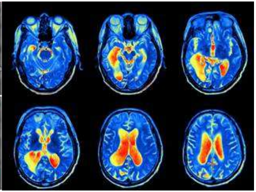
In many applications $\rightsquigarrow$ interest is in learning about unknowns from observed (noisy) data











General context: Bayesian inference

Bayesian $\Rightarrow$ Instead of just a pointwise estimation of the unknowns

$$\widehat{x}_{1:n}$$

we are interested in its posterior probability density function (pdf)

$$\pi_n(x_{1:n}) \equiv p(x_{1:n}|y_{1:n}) \propto p(x_{1:n})p(y_{1:n}|x_{1:n})$$

thus characterizing all the uncertainty in the model under study.

General context: Bayesian inference

Bayesian $\Rightarrow$ Instead of just a pointwise estimation of the unknowns

$$\widehat{x}_{1:n}$$

we are interested in its posterior probability density function (pdf)

$$\pi_n(x_{1:n}) \equiv p(x_{1:n}|y_{1:n}) \propto p(x_{1:n})p(y_{1:n}|x_{1:n})$$

thus characterizing all the uncertainty in the model under study.

Unfortunately, in most cases, this posterior pdf is intractable

$\Rightarrow$ all quantities of interest such as

$$\Theta_n = \mathbb{E}_{\pi_n}[\varphi(X_n)] = \int \varphi(x_n)\pi_n(x_n)dx_n$$

cannot be computed and must be approximated

Sequential Monte-Carlo algorithms

Outline

Importance Sampling

Let us define the target distribution of interest which is known up to a normalizing constant Z:

$$\pi(x) = \frac{\gamma(x)}{Z}$$

Importance Sampling (IS) identity

For any distribution q such that $supp(\pi) \subset supp(q)$

$$\mathbb{E}_\pi[h(X)] = \int h(x) \frac{\pi(x)}{q(x)} q(x) dx$$

$q(\cdot)$ is called *importance (or proposal / instrumental) distribution*

$q(\cdot)$ can be chosen arbitrarily, in particular easy to sample from.

Importance Sampling

Let us define the target distribution of interest which is known up to a normalizing constant Z:

$$\pi(x) = \frac{\gamma(x)}{Z}$$

Importance Sampling (IS) identity

For any distribution q such that $supp(\pi) \subset supp(q)$

$$\begin{aligned}\mathbb{E}_\pi[h(X)] &= \int h(x) \frac{\pi(x)}{q(x)} q(x) dx \\ &= \int h(x) w(x) q(x) dx = \mathbb{E}_q[h(X) w(x)]\end{aligned}$$

$q(\cdot)$ is called *importance (or proposal / instrumental) distribution*
$w(x) = \pi(x)/q(x)$ is called *importance weight*.

$q(\cdot)$ can be chosen arbitrarily, in particular easy to sample from.

Importance Sampling

Target distribution of interest: $\pi(x) = \frac{\gamma(x)}{Z}$

Importance Sampling (IS) identity

For any distribution q such that $supp(\pi) \subset supp(q)$

$$\mathbb{E}_\pi[h(X)] = \int h(x)w(x)q(x)dx \quad \text{with } w(x) = \pi(x)/q(x)$$

- Draw independently N_p samples from $q(\cdot)$

$$\text{for } j = 1, \dots, N_p : \; X^j \overset{iid}{\sim} q(\cdot)$$

- Plugging this expression in the IS identity, we obtain [by the Law of Large numbers]:

$$\frac{1}{N}\sum_{i=1}^{N} w(X^i)h(X^i) \xrightarrow[N \to \infty]{a.s.} \mathbb{E}_\pi[h(X)]$$

or self-normalized version (used when Z is unknown)

$$\sum_{i=1}^{N} \frac{w(X^i)}{\sum_{j=1}^{N} w(X^j)} h(X^i) \xrightarrow[N \to \infty]{a.s.} \mathbb{E}_\pi[h(X)]$$

$$\text{with } w(X^i) = \frac{\gamma(X^i)}{q(X^i)}$$

Outline

General SMC methodology

Aim: Approximate a sequence of target pdf of increasing dimension

$$\left\{\pi_n(x_{1:n})\right\}_{n\geq 1}$$

i.e. the dimension of its support forms an increasing sequence with n.

- In practice: π_n only known up to a normalizing constant,

$$\pi_n(x_{1:n}) = \frac{\gamma_n(x_{1:n})}{Z_n}$$

General SMC methodology

Aim: Approximate a sequence of target pdf of increasing dimension

$$\left\{\pi_n(x_{1:n})\right\}_{n\geq 1}$$

i.e. the dimension of its support forms an increasing sequence with n.

- In practice: π_n only known up to a normalizing constant,

$$\pi_n(x_{1:n}) = \frac{\gamma_n(x_{1:n})}{Z_n}$$

- Originally developed for filtering in hidden Markov model (HMM) with

$$\gamma_n(x_{1:n}) = p(x_{1:n}, y_{1:n})$$

BUT can be used for other sequence of target pdf.

How does SMC work?

- Sequence of *importance sampling* (IS) steps, where at each step n
  - the target distribution is $\pi_n(x_{1:n}) = \gamma_n(x_{1:n})/Z_n$
  - the importance distribution is $q_n(x_{1:n}) = q_1(x_1)\prod_{k=2}^{n} q_k(x_k|x_{1:k-1})$

How does SMC work?

- Sequence of *importance sampling* (IS) steps, where at each step n
  - the target distribution is $\pi_n(x_{1:n}) = \gamma_n(x_{1:n})/Z_n$
  - the importance distribution is $q_n(x_{1:n}) = q_1(x_1) \prod_{k=2}^{n} q_k(x_k|x_{1:k-1})$

Procedure at time n and $\forall j = 1, \dots, N$ [Gordon et al, 1993]

- **Sampling** - Propagate each trajectory: $\widetilde{X}_n^j \sim q_n(x_n|X_{1:n-1}^j)$

How does SMC work?

- Sequence of *importance sampling* (IS) steps, where at each step n
  - the target distribution is $\pi_n(x_{1:n}) = \gamma_n(x_{1:n})/Z_n$
  - the importance distribution is $q_n(x_{1:n}) = q_1(x_1)\prod_{k=2}^n q_k(x_k|x_{1:k-1})$

Procedure at time n and $\forall j = 1, \ldots, N$ [Gordon et al, 1993]

- **Sampling** - Propagate each trajectory: $\widetilde{X}_n^j \sim q_n(x_n|X_{1:n-1}^j)$
- **Weighting** - Compute each importance weight:

$$W_n^j \propto W(X_{1:n-1}^j, \widetilde{X}_n^j) = \frac{\gamma_n(X_{1:n-1}^j, \widetilde{X}_n^j)}{q_n(X_{1:n-1}^j, \widetilde{X}_n^j)} = \underbrace{\frac{\gamma_{n-1}(X_{1:n-1}^j)}{q_{n-1}(X_{1:n-1}^j)}}_{\substack{W(X_{1:n-1}^j) \\ \text{Previous weight}}} \underbrace{\frac{\gamma_n(X_{1:n-1}^j, \widetilde{X}_n^j)}{\gamma_{n-1}(X_{1:n-1}^j) q_n(\widetilde{X}_n^j|X_{1:n-1}^j)}}_{\substack{\widetilde{w}(X_{1:n-1}^j, \widetilde{X}_n^j) \\ \text{Incremental weight}}}$$

$\Rightarrow$ for filtering problem in HMM with $\gamma_n(x_{1:n}) = p(x_{1:n}, y_{1:n})$, we simply have

$$\widetilde{w}(X_{1:n-1}^j, \widetilde{X}_n^j) = \frac{g_n(y_n|\widetilde{X}_n^j) f_n(\widetilde{X}_n^j|X_{n-1}^j)}{q_n(\widetilde{X}_n^j|X_{1:n-1}^j)}.$$

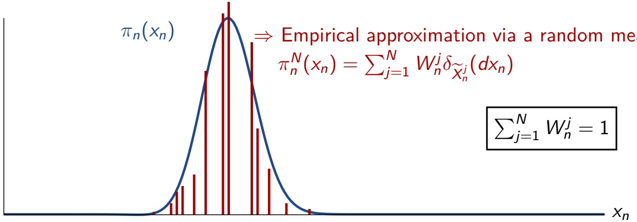
How does SMC work?

- Sequence of *importance sampling* (IS) steps, where at each step n
  - the target distribution is $\pi_n(x_{1:n}) = \gamma_n(x_{1:n})/Z_n$
  - the importance distribution is $q_n(x_{1:n}) = q_1(x_1)\prod_{k=2}^n q_k(x_k|x_{1:k-1})$

Procedure at time n and $\forall j = 1, \ldots, N$ [Gordon et al, 1993]

- **Sampling** - Propagate each trajectory: $\widetilde{X}_n^j \sim q_n(x_n|X_{1:n-1}^j)$
- **Weighting** - Compute each importance weight:

$$W_n^j \propto W(X_{1:n-1}^j, \widetilde{X}_n^j) = \frac{\gamma_n(X_{1:n-1}^j, \widetilde{X}_n^j)}{q_n(X_{1:n-1}^j, \widetilde{X}_n^j)} = \underbrace{\frac{\gamma_{n-1}(X_{1:n-1}^j)}{q_{n-1}(X_{1:n-1}^j)}}_{\substack{W(X_{1:n-1}^j) \\ \text{Previous weight}}} \underbrace{\frac{\gamma_n(X_{1:n-1}^j, \widetilde{X}_n^j)}{\gamma_{n-1}(X_{1:n-1}^j)q_n(\widetilde{X}_n^j|X_{1:n-1}^j)}}_{\substack{\widetilde{w}(X_{1:n-1}^j, \widetilde{X}_n^j) \\ \text{Incremental weight}}}$$



$\Rightarrow$ Empirical approximation via a random measure $\{W_n^j, \widetilde{X}_n^j\}_{j=1}^N$

$\pi_n^N(x_n) = \sum_{j=1}^N W_n^j \delta_{\widetilde{X}_n^j}(dx_n)$

$$\boxed{\sum_{j=1}^N W_n^j = 1}$$

How does SMC work?

- Sequence of *importance sampling* (IS) steps, where at each step n
  - the target distribution is $\pi_n(x_{1:n}) = \gamma_n(x_{1:n})/Z_n$
  - the importance distribution is $q_n(x_{1:n}) = q_1(x_1)\prod_{k=2}^{n} q_k(x_k|x_{1:k-1})$

Procedure at time n and $\forall j = 1, \dots, N$ [Gordon et al, 1993]

- **Sampling** - Propagate each trajectory: $\widetilde{X}_n^j \sim q_n(x_n|X_{1:n-1}^j)$
- **Weighting** - Compute each importance weight:

$$W_n^j \propto W(X_{1:n-1}^j, \widetilde{X}_n^j) = W(X_{1:n-1}^j)\widetilde{w}(X_{1:n-1}^j, \widetilde{X}_n^j)$$

$$\Theta_n = \mathbb{E}_{\pi_n}[\varphi(X_n)] \Rightarrow \boxed{\widehat{\Theta}_n^{SIS,N} = \sum_{j=1}^{N} W_n^j \varphi(\widetilde{X}_n^j)}$$

Sequential Importance Sampling

Sequential Importance Sampling Algorithm

At time 1: for $j = 1, \dots, N_p$, sample $X_1^j \sim q_1(x_1)$ and set $w_1^j = \frac{\gamma_1(X_1^j)}{q_1(X_1^j)}$

for time $k > 1$ **do**

  for $j = 1, \dots, N$ **do**

    Sample $\widetilde{X}_n^j \sim q_n(\cdot|X_{n-1}^j)$

    Compute Importance weight $W_n^j \propto W_{n-1}^j \frac{\gamma_n(X_{1:n-1}^j, \widetilde{X}_n^j)}{\gamma_{n-1}(X_{1:n-1}^j) q_n(\widetilde{X}_n^j | X_{1:n-1}^j)}$

    Set $X_n^j = \widetilde{X}_n^j$

  end for

  Output Approximations :

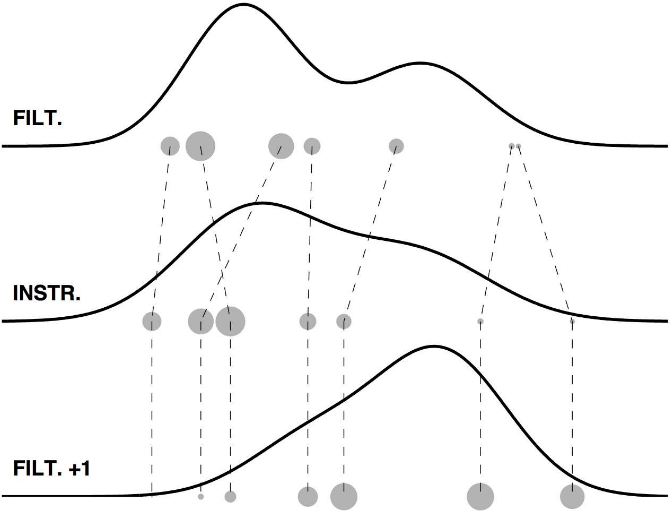
  Target pdf: $\widehat{\pi}_n(x_{1:n}) = \sum_{i=1}^N W_n^j \delta_{X_{1:n}^i}(dx_{1:n})$

  $\Theta_n = \mathbb{E}_{\pi_n}[\varphi(X_n)] \approx \widehat{\Theta}_n^{SIS,N} = \sum_{j=1}^N W_n^j \varphi(\widetilde{X}_n^j)$

  Normalizing constant: $\widehat{Z}_n = \sum_{j=1}^N W_n^i$

end for

Sequential Importance Sampling



One step of the SIS algorithm with just seven particles.

SIS : Choice of the proposal distribution

The so-called "optimal" choice of $q_n(\cdot)$ (for filtering in HMM) that minimizes the variance of the importance weights, consists in setting

$$q_n(x_n|x_{n-1}) = \frac{g_n(y_n|x_n)f_n(x_n|x_{n-1})}{\int g_n(y_n|x_n)f_n(x_n|x_{n-1})dx_n}$$

Consequently the weights does not depend on current state value but only on the previous one ($\mathbf{x}_{n-1}$), i.e. :

$$W_n^j \propto W_{n-1}^j \int g_n(y_n|x_n)f_n(x_n|x_{n-1}^j)dx_n$$

SIS : Choice of the proposal distribution

The so-called "optimal" choice of $q_n(\cdot)$ (for filtering in HMM) that minimizes the variance of the importance weights, consists in setting

$$q_n(x_n|x_{n-1}) = \frac{g_n(y_n|x_n)f_n(x_n|x_{n-1})}{\int g_n(y_n|x_n)f_n(x_n|x_{n-1})dx_n}$$

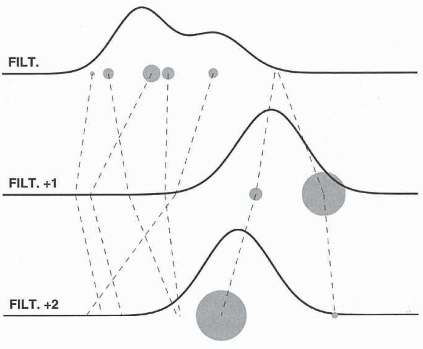
Consequently the weights does not depend on current state value but only on the previous one ($\mathbf{x}_{n-1}$), i.e. :

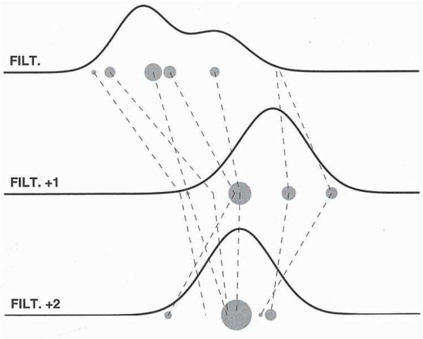
$$W_n^j \propto W_{n-1}^j \int g_n(y_n|x_n)f_n(x_n|x_{n-1}^j)dx_n$$

This is however usually not feasible and common choices include :

- the prior $q_n(x_n|x_{n-1}) = f_n(x_n|x_{n-1})$ (and then $W_n^j \propto W_{n-1}^j g_n(y_n|X_n^j)$),
- approximations (sometimes heuristic) to the optimal one (moment matching, use of EKF or UKF, ...),
- tuning parameters of $q_n(\cdot)$ so as to maximize some criterions: effective sample size, entropy, ...

SIS : Choice of the proposal distribution

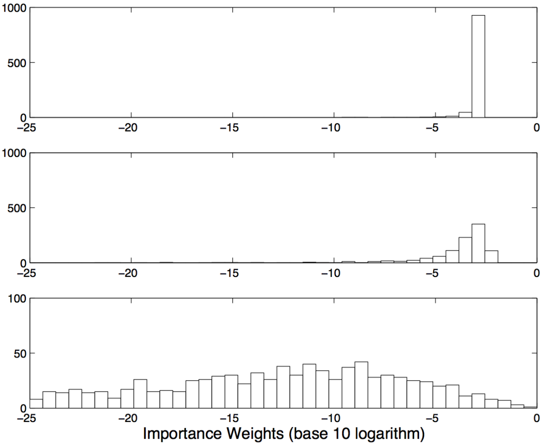




SIS with the of the prior (left) and the optimal (right) distribution as proposal.

⇒ Choice of this proposal distribution is an important step when one want to design an efficient SIS algorithm.

SIS: Weight Degeneracy



Histograms of the base 10 logarithm of the normalized weights for $t = 1$ *(top),* $t = 50$ *(middle) and* $t = 100$ *(bottom) for a simple stochastic volatility model.*

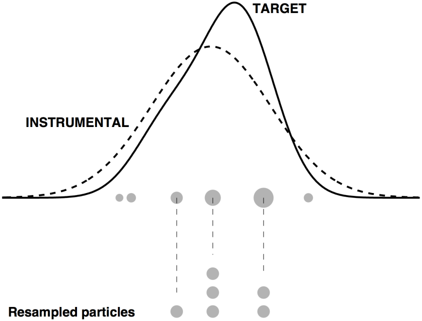
The algorithm performance collapse as time (n) increases... After a few time steps, only a very small number of particles have non negligible weigths !!

SIS: Resampling Step

Problem: After a few time steps, only a very small number of particles have non negligible weigths !

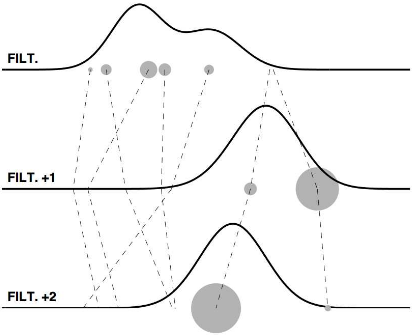
Solution: Replicate particles with large weights and eliminate those with small weights to prevent the problems we saw with SIS (at the price of a, usually moderate, increase in variance).
↪ Use random resampling techniques by taking importance weights: multinomial, residual, ...

⇒ Sequential Importance Resampling



Sequential Importance Sampling Resampling

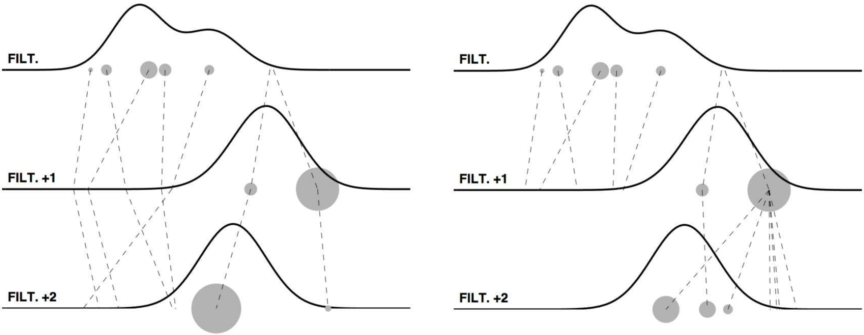
SIS (left) versus SIR (right)



The resampling step is necessary to ensure the long-term stability of the filtering algorithm.

$\hookrightarrow$ Maintain a reasonable number of contributing particles at all times.

SIS (left) versus SIR (right)



The resampling step is necessary to ensure the long-term stability of the filtering algorithm.

$\hookrightarrow$ Maintain a reasonable number of contributing particles at all times.

BUT it reduces the number of distinct samples $\Rightarrow$ *sample impoverishment*

SMC technique: summary

- Sequence of *importance sampling* (IS) steps, where at each step n
  - the target distribution is $\pi_n(x_{1:n}) = \gamma_n(x_{1:n})/Z_n$
  - the importance distribution is $q_n(x_{1:n}) = q_1(x_1)\prod_{k=2}^{n} q_k(x_k|x_{1:k-1})$

Procedure at time n and $\forall j = 1, \ldots, N$ [Gordon et al, 1993]

- **Sampling** - Propagate each trajectory: $\widetilde{X}_n^j \sim q_n(x_n|X_{1:n-1}^j)$
- **Weighting** - Compute each importance weight:

$$W_n^j \propto W(X_{1:n-1}^j, \widetilde{X}_n^j) = W(X_{1:n-1}^j)\widetilde{w}(X_{1:n-1}^j, \widetilde{X}_n^j)$$

$$\Theta_n = \mathbb{E}_{\pi_n}[\varphi(X_n)] \Rightarrow \boxed{\widehat{\Theta}_n^{SIS,N} = \sum_{j=1}^{N} W_k^j \varphi(\widetilde{X}_n^j)}$$

- **Resampling** (optional but necessary to avoid *weight degeneracy*): Sample N times from the random measure $\{W_n^j, X_{1:n-1}^j, \widetilde{X}_n^j\}_{j=1}^N$ to obtain $\{W_n^j = \frac{1}{N}, X_{1:n}^j\}_{j=1}^N$

$$\boxed{\widehat{\Theta}_n^{SIR,N} = \frac{1}{N}\sum_{j=1}^{N} \varphi(X_n^j)}$$

Outline

Effect of the (optional) multinomial resampling step:

Fights against weight degeneracy but no local benefits:

- dependency among the resampled points $\Rightarrow$ support shrinkage
- $\text{var}(\widehat{\Theta}_n^{SIR,N}) \geq \text{var}(\widehat{\Theta}_n^{SIS,N})$

Effect of the (optional) multinomial resampling step:

Fights against weight degeneracy but no local benefits:

- dependency among the resampled points $\Rightarrow$ support shrinkage
- $\text{var}(\widehat{\Theta}_n^{SIR,N}) \geq \text{var}(\widehat{\Theta}_n^{SIS,N})$

Key idea: *Ph.D thesis of R. Lamberti - [Lamberti et al, IEEE TSP 2017]*

Revisit the complete scheme
(sampling, weighting and resampling)

$\downarrow$

Benefits of the resampling mechanism
without local impoverishment of the resulting MC approx.

Outline

Characterization of sampling / weighting / resampling step

Proposition 1

Given $\{X^i_{0:n-1}\}_{i=1}^N$, the resampled particles X^j_n are identically distributed **(but dependent)** according to a pdf $\widetilde{q}^N_n(x)$, with

$$\widetilde{q}^N_n(x) = \sum_{i=1}^N q^i_n(x) h^i_n(x) \quad (1)$$

where

$$\begin{aligned} q^i_n(x) &= q(x|X^i_{0:n-1}) \\ h^i_n(x) &= \int \int \frac{\frac{\pi^i_n(x)}{q^i_n(x)}}{\frac{\pi^i_n(x)}{q^i_n(x)} + \sum_{l \neq i} \frac{\pi^l_n(x^l)}{q^l_n(x^l)}} \prod_{l \neq i} q^l_n(x^l) dx^l \end{aligned} \quad (2)$$

This dependency results in support shrinkage since, by construction, an intermediate sample can be resampled several times.

Characterization of sampling / weighting / resampling step

Proposition 1

Given $\{X^i_{0:n-1}\}_{i=1}^N$, the resampled particles X^j_n are identically distributed **(but dependent)** according to a pdf $\widetilde{q}^N_n(x)$, with

$$\widetilde{q}^N_n(x) = \sum_{i=1}^N q^i_n(x) h^i_n(x) \quad (1)$$

where

$$q^i_n(x) = q(x|X^i_{0:n-1})$$
$$h^i_n(x) = \int\int \frac{\frac{\pi^i_n(x)}{q^i_n(x)}}{\frac{\pi^i_n(x)}{q^i_n(x)} + \sum_{l\neq i} \frac{\pi^l_n(x^l)}{q^l_n(x^l)}} \prod_{l\neq i} q^l_n(x^l) dx^l \quad (2)$$

Proposition ⇒ Scheme to produce independent samples from $\widetilde{q}^N_n(x)$

Proposed independent resampling

Method to sample N **independent** particles from $\widetilde{q}_n^N(x)$

$\Rightarrow$ **weighting & resampling as compound importance distrib.**

Proposed independent resampling

Method to sample N **independent** particles from $\widetilde{q}_n^N(x)$

$\Rightarrow$ **weighting & resampling as compound importance distrib.**

Example with $N = 8$: we obtain 8 particles drawn i.i.d. from $\widetilde{q}_n^8(x)$

Proposed independent resampling

Method to sample N **independent** particles from $\widetilde{q}_n^N(x)$

$\Rightarrow$ **weighting & resampling as compound importance distrib.**

Example with $N = 8$: we obtain 8 particles drawn i.i.d. from $\widetilde{q}_n^8(x)$

Proposed independent resampling

Method to sample N **independent** particles from $\widetilde{q}_n^N(x)$

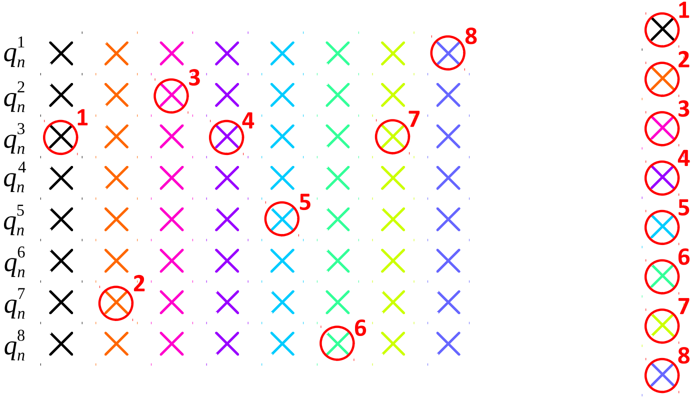
$\Rightarrow$ **weighting & resampling as compound importance distrib.**

Example with $N = 8$: we obtain 8 particles drawn i.i.d. from $\widetilde{q}_n^8(x)$

Proposed independent resampling

Method to sample N **independent** particles from $\widetilde{q}_n^N(x)$

$\Rightarrow$ **weighting & resampling as compound importance distrib.**



Example with $N = 8$: we obtain 8 particles drawn i.i.d. from $\widetilde{q}_n^8(x)$

Proposed independent resampling

Method to sample N **independent** particles from $\widetilde{q}_n^N(x)$

$\Rightarrow$ **weighting & resampling as compound importance distrib.**

- Sampling N independent particles from $\widetilde{q}_n^N(x)$
- Weighting - 2 solutions:
  1. Equal weights: $W_n^i = 1/N$ - what is actually done after classical resampling
     $\Rightarrow$ Algorithm I-SIR

Proposed independent resampling

Method to sample N **independent** particles from $\widetilde{q}_n^N(x)$

$\Rightarrow$ **weighting & resampling as compound importance distrib.**

- Sampling N independent particles from $\widetilde{q}_n^N(x)$
- Weighting - 2 solutions:
  1. Equal weights: $W_n^i = 1/N$ - what is actually done after classical resampling
     $\Rightarrow$ Algorithm I-SIR
  2. IS principle: $W_n^i \propto \pi_n^i(X_n^i) \Big/ \widetilde{q}_n^N(X_n^i)$
     $\Rightarrow$ Algorithm I-SIR-w

     However $\widetilde{q}_n^N(X_n^i) = \sum_{i=1}^N q_n^i(x) h_n^i(x)$ must be approximated

     since $h_n^i(x) = \int \int \frac{\frac{\pi_n^i(x)}{q_n^i(x)}}{\frac{\pi_n^i(x)}{q_n^i(x)} + \sum_{l \neq i} \frac{\pi_n^l(x^l)}{q_n^l(x^l)}} \prod_{l \neq i} q_n^l(x^l) dx^l$ cannot be evaluated

     $\Rightarrow$ h_n^i approximated by the N^2 samples

Proposed independent resampling

Contributions $\mapsto$ Theoretical study of $\widehat{\Theta}_n^{\mathrm{I-SIR},N}$

- Mean/Variance for finite N (number of samples)

$$\mathbb{E}(\widehat{\Theta}_n^{\mathrm{I-SIR},N}|\{X_{0:n-1}^i\}_{i=1}^N) = \mathbb{E}(\widehat{\Theta}_n^{\mathrm{SIR},N}|\{X_{0:n-1}^i\}_{i=1}^N) = \mathbb{E}(\widehat{\Theta}_n^{\mathrm{SIS},N}|\{X_{0:n-1}^i\}_{i=1}^N),$$

Proposed independent resampling

Contributions $\mapsto$ Theoretical study of $\widehat{\Theta}_n^{\mathrm{I-SIR},N}$

- Mean/Variance for finite N (number of samples)

$$\mathbb{E}(\widehat{\Theta}_n^{\mathrm{I-SIR},N}|\{X_{0:n-1}^i\}_{i=1}^N) = \mathbb{E}(\widehat{\Theta}_n^{\mathrm{SIR},N}|\{X_{0:n-1}^i\}_{i=1}^N) = \mathbb{E}(\widehat{\Theta}_n^{\mathrm{SIS},N}|\{X_{0:n-1}^i\}_{i=1}^N),$$

$$\mathrm{var}(\widehat{\Theta}_n^{\mathrm{I-SIR},N}|\{X_{0:n-1}^i\}_{i=1}^N) = \mathrm{var}(\widehat{\Theta}_n^{\mathrm{SIR},N}|\{X_{0:n-1}^i\}_{i=1}^N) - \frac{N-1}{N}\mathrm{var}(\widehat{\Theta}_n^{\mathrm{SIS},N}|\{X_{0:n-1}^i\}_{i=1}^N).$$

  - The proposed algorithm ($\widehat{\Theta}_n^{\mathrm{I-SIR}}$) would outperform the classical SIR estimate
  - Gain depends on $\mathrm{var}(\widehat{\Theta}_n^{\mathrm{SIS},N}|\{X_{0:n-1}^i\}_{i=1}^N)$

Proposed independent resampling

Contributions $\mapsto$ Theoretical study of $\widehat{\Theta}_n^{\mathrm{I-SIR},N}$

- CLT for a single step

$$\sqrt{N}(\widehat{\Theta}^{\mathrm{I-SIR},N} - \Theta) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \underbrace{\mathrm{var}_\pi(\varphi(x))}_{\sigma_\infty^{2,\mathrm{I-SIR}}(q)}\right).$$

for comparison:

$$\sigma_\infty^{2,\mathrm{SIS}}(q) = \mathbb{E}_q\left(\frac{\pi^2(x)}{q^2(x)}(\varphi(x) - \Theta)^2\right),$$

$$\sigma_\infty^{2,\mathrm{SIR}}(q) = \mathrm{var}_\pi(\varphi(x)) + \mathbb{E}_q\left(\frac{\pi^2(x)}{q^2(x)}(\varphi(x) - \Theta)^2\right)$$

  - Asymp. variance: $\sigma_\infty^{2,\mathrm{I-SIR}}(q) \leq \sigma_\infty^{2,\mathrm{SIR}}(q)$
  - $\sigma_\infty^{2,\mathrm{I-SIR}}(q)$ no longer depends on the proposal distribution q

Proposed independent resampling

Contributions $\mapsto$ Theoretical study of $\widehat{\Theta}_n^{\mathrm{I-SIR},N}$

- CLT for a single step

$$\sqrt{N}(\widehat{\Theta}^{\mathrm{I-SIR},N} - \Theta) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \underbrace{\mathrm{var}_\pi(\varphi(x))}_{\sigma_\infty^{2,\mathrm{I-SIR}}(q)}\right).$$

for comparison:

$$\sigma_\infty^{2,\mathrm{SIS}}(q) = \mathbb{E}_q\left(\frac{\pi^2(x)}{q^2(x)}(\varphi(x)-\Theta)^2\right),$$

$$\sigma_\infty^{2,\mathrm{SIR}}(q) = \mathrm{var}_\pi(\varphi(x)) + \mathbb{E}_q\left(\frac{\pi^2(x)}{q^2(x)}(\varphi(x)-\Theta)^2\right)$$

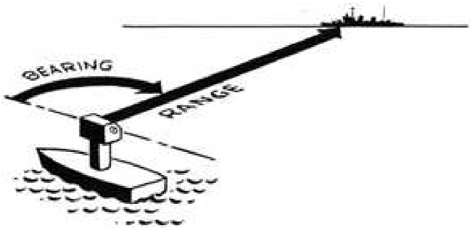
  - Asymp. variance: $\sigma_\infty^{2,\mathrm{I-SIR}}(q) \leq \sigma_\infty^{2,\mathrm{SIR}}(q)$
  - $\sigma_\infty^{2,\mathrm{I-SIR}}(q)$ no longer depends on the proposal distribution q
  - Difficult to compare $\sigma_\infty^{2,\mathrm{I-SIR}}(q)$ and $\sigma_\infty^{2,\mathrm{SIS}}(q)$

Outline

Numerical simulations - Tracking

- State of interest $x_n = [p_{x,n}, \dot{p}_{x,n}, p_{y,n}, \dot{p}_{y,n}]^T$ position and velocity of a target
- Tracking scenario with range-bearing measurements $\Rightarrow$ likelihood pdf:

$$g_n(y_n|x_n) = \mathcal{N}\left(y_n; \begin{bmatrix} \sqrt{p_{x,n}^2 + p_{y,n}^2} \\ \arctan \frac{p_{y,n}}{p_{x,n}} \end{bmatrix}, R\right) \text{ with } R = \begin{pmatrix} \sigma_\rho^2 & 0 \\ 0 & \sigma_\theta^2 \end{pmatrix}$$



Numerical simulations - Tracking

- State of interest $x_n = [p_{x,n}, \dot{p}_{x,n}, p_{y,n}, \dot{p}_{y,n}]^T$ position and velocity of a target
- Tracking scenario with range-bearing measurements $\Rightarrow$ likelihood pdf:

$$g_n(y_n|x_n) = \mathcal{N}\left(y_n; \begin{bmatrix} \sqrt{p_{x,n}^2 + p_{y,n}^2} \\ \arctan \frac{p_{y,n}}{p_{x,n}} \end{bmatrix}, R\right) \text{ with } R = \begin{pmatrix} \sigma_\rho^2 & 0 \\ 0 & \sigma_\theta^2 \end{pmatrix}$$

- Prior knowledge on the dynamics of the unknown state:

Near Constant Velocity (NCV) model

$$f_n(x_n|x_{n-1}) = \mathcal{N}(x_n; Fx_{n-1}, Q)$$

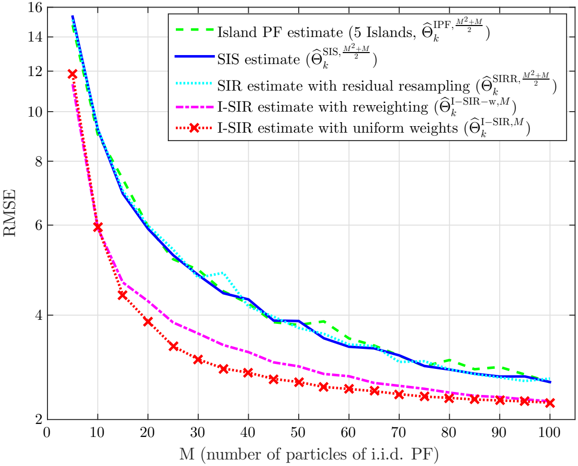
with

$$F = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \text{ and } Q = \sigma_Q^2 \begin{pmatrix} \frac{1}{3} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & 1 \end{pmatrix}$$

Numerical simulations - Tracking

Target tracking scenario: range-bearing measurements ($\sigma_Q = \sqrt{10}$, $\sigma_\rho = 0.05$, $\sigma_\theta = \frac{\pi}{3600}$)

Comparison with other SMC algorithms in which the number of particles is chosen so that they have the same computational cost



For RMSE=2.7

$N_{\mathrm{I-SIR}} = 50$

vs

$N_{\mathrm{SIS}} = 5050$

Proposed independent resampling

Summary of the proposed I-SIR

- No support degeneracy: better particle diversity for the next iteration
- Gain in terms of variance of the resulting estimator

Proposed independent resampling

Summary of the proposed I-SIR

- No support degeneracy: better particle diversity for the next iteration
- Gain in terms of variance of the resulting estimator

but

- N^2 sampling & weighting steps: higher computational cost than the classical PF if $N_{SIS} = N_{\mathrm{I}-SIR}$.

Proposed independent resampling

Summary of the proposed I-SIR

- No support degeneracy: better particle diversity for the next iteration
- Gain in terms of variance of the resulting estimator

but

- N^2 sampling & weighting steps: higher computational cost than the classical PF if $N_{SIS} = N_{\text{I}-SIR}$.

  However

  - Resampling is not necessarily needed at each iteration
  - Independent resampling can be parallelized
  - In some cases [as in before], performs better even when $N_{\text{I}-SIR}^2 + N_{\text{I}-SIR} = 2N_{SIS}$

Proposed independent resampling

Summary of the proposed I-SIR

- No support degeneracy: better particle diversity for the next iteration
- Gain in terms of variance of the resulting estimator

but

- N^2 sampling & weighting steps: higher computational cost than the classical PF if $N_{SIS} = N_{\mathrm{I}-SIR}$.
  However
  - Resampling is not necessarily needed at each iteration
  - Independent resampling can be parallelized
  - In some cases [as in before], performs better even when $N^2_{\mathrm{I}-SIR} + N_{\mathrm{I}-SIR} = 2N_{SIS}$

Can we propose a general framework that will include classical and independent as a special case ?

Outline

Classical vs Independent

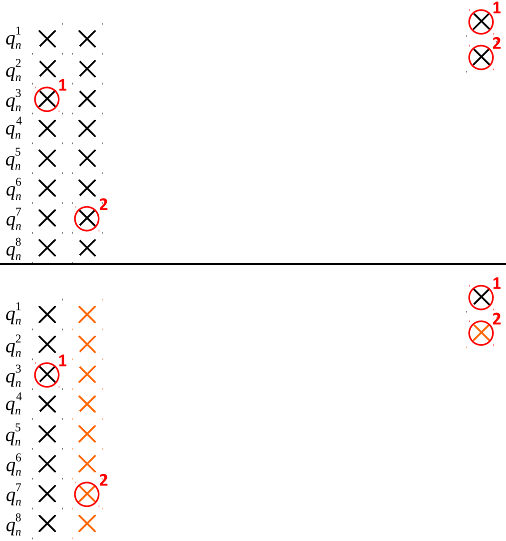
Classical
SIR

Independent
I-SIR

Classical vs Independent

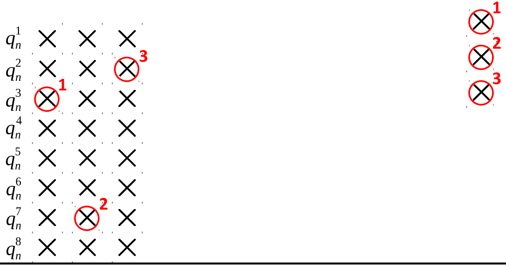
Classical
SIR

Independent
I-SIR



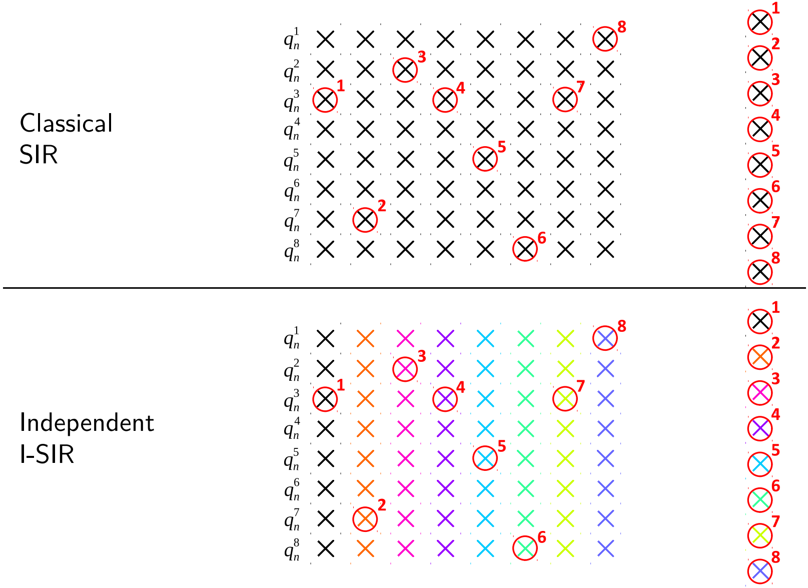
Classical vs Independent

Classical
SIR



Independent
I-SIR

Classical vs Independent



Idea $\mapsto$ Regenerate only $0 \leq k \leq N = 8$ samples per iterations

The semi-independent algorithm

Independent
I-SIR

Semi-Independent
SR

Here - $k = 3$ uniformly chosen samples are redrawn from the previous population

The semi-independent algorithm

Here - $k = 3$ uniformly chosen samples are redrawn from the previous population

The semi-independent algorithm

Independent
I-SIR

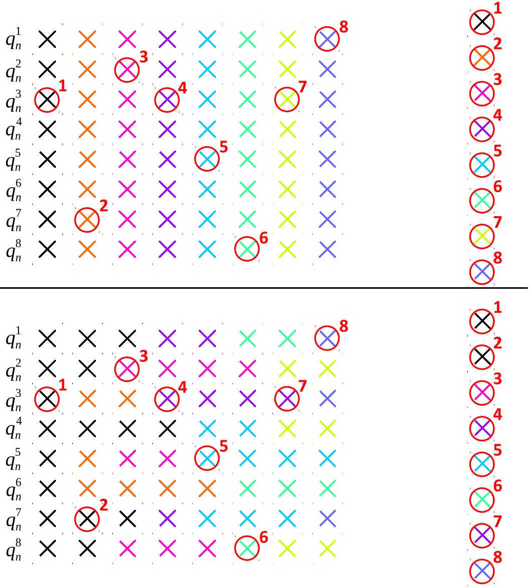
Semi-Independent
SR

Here - $k = 3$ uniformly chosen samples are redrawn from the previous population

The semi-independent algorithm

Independent
I-SIR

Semi-Independent
SR



Remark: Not anymore parallelized when $0 < k < N$!

The semi-independent algorithm

Semi-Independent SR

Semi-Independent NSSR

NSSR is a fully parallelized version of this semi-independent resampling

The semi-independent algorithm

Semi-Independent SR

Semi-Independent NSSR

NSSR is a fully parallelized version of this semi-independent resampling

The semi-independent algorithm

Semi-Independent SR

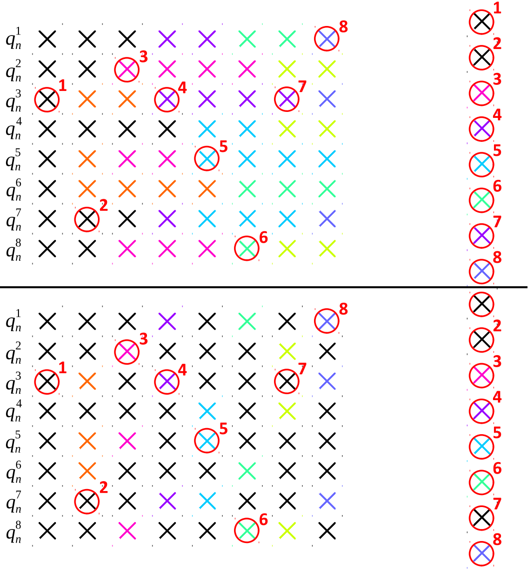
Semi-Independent NSSR

NSSR is a fully parallelized version of this semi-independent resampling

The semi-independent algorithm

Semi-Independent SR

Semi-Independent NSSR



NSSR is a fully parallelized version of this semi-independent resampling

Semi-independent algorithm - Theoretical analysis

- **Generalization** of classical SIR and the previous independent I-SIR
  - When $k = 0$ regenerated samples $\Rightarrow$ SR=NSSR=SIR
  - When $k = N$ regenerated samples $\Rightarrow$ SR=NSSR=I-SIR

Semi-independent algorithm - Theoretical analysis

- **Generalization** of classical SIR and the previous independent I-SIR
  - When $k = 0$ regenerated samples $\Rightarrow$ SR=NSSR=SIR
  - When $k = N$ regenerated samples $\Rightarrow$ SR=NSSR=I-SIR
- **Statistical properties:** Given the previous set of particles $\{X_{0:n-1}^i\}_{i=1}^N$, for all k, $0 \leq k \leq N$, we have:

$$\mathrm{E}(\widehat{\Theta}_t^{\mathrm{NSSR},N,k}) = \mathrm{E}(\widehat{\Theta}_t^{\mathrm{SR},N,k}) = \mathrm{E}(\widehat{\Theta}_t^{\mathrm{I-SIR},N}) = \mathrm{E}(\widehat{\Theta}_t^{\mathrm{SIR},N})$$

Each sample X_n^i are marginally drawn from the same distribution $\widetilde{q}_n^N(x)$

Semi-independent algorithm - Theoretical analysis

- **Generalization** of classical SIR and the previous independent I-SIR
  - When $k = 0$ regenerated samples $\Rightarrow$ SR=NSSR=SIR
  - When $k = N$ regenerated samples $\Rightarrow$ SR=NSSR=I-SIR
- **Statistical properties:** Given the previous set of particles $\{X^i_{0:n-1}\}_{i=1}^N$, for all k, $0 \leq k \leq N$, we have:

$$\mathrm{E}(\widehat{\Theta}_t^{\mathrm{NSSR},N,k}) = \mathrm{E}(\widehat{\Theta}_t^{\mathrm{SR},N,k}) = \mathrm{E}(\widehat{\Theta}_t^{\mathrm{I-SIR},N}) = \mathrm{E}(\widehat{\Theta}_t^{\mathrm{SIR},N})$$

$$\mathrm{var}(\widehat{\Theta}_t^{\mathrm{I-SIR},N}) \leq \mathrm{var}(\widehat{\Theta}_t^{\mathrm{SR},N,k}) \leq \mathrm{var}(\widehat{\Theta}_t^{\mathrm{NSSR},N,k}) \leq \mathrm{var}(\widehat{\Theta}_t^{\mathrm{SIR},N})$$

Semi-independent algorithm - Theoretical analysis

- **Generalization** of classical SIR and the previous independent I-SIR
  - When $k = 0$ regenerated samples $\Rightarrow$ SR=NSSR=SIR
  - When $k = N$ regenerated samples $\Rightarrow$ SR=NSSR=I-SIR
- **Statistical properties:** Given the previous set of particles $\{X^i_{0:n-1}\}_{i=1}^N$, for all k, $0 \leq k \leq N$, we have:

$$\mathrm{E}(\widehat{\Theta}_t^{\mathrm{NSSR},N,k}) = \mathrm{E}(\widehat{\Theta}_t^{\mathrm{SR},N,k}) = \mathrm{E}(\widehat{\Theta}_t^{\mathrm{I-SIR},N}) = \mathrm{E}(\widehat{\Theta}_t^{\mathrm{SIR},N})$$

$$\mathrm{var}(\widehat{\Theta}_t^{\mathrm{I-SIR},N}) \leq \mathrm{var}(\widehat{\Theta}_t^{\mathrm{SR},N,k}) \leq \mathrm{var}(\widehat{\Theta}_t^{\mathrm{NSSR},N,k}) \leq \mathrm{var}(\widehat{\Theta}_t^{\mathrm{SIR},N})$$

$$\mathrm{var}(\widehat{\Theta}_t^{\mathrm{SR},N,k}) \leq \mathrm{var}(\widehat{\Theta}_t^{\mathrm{SR},N,k-1})$$

Semi-independent algorithm - Theoretical analysis

- **Generalization** of classical SIR and the previous independent I-SIR
  - When $k = 0$ regenerated samples $\Rightarrow$ SR=NSSR=SIR
  - When $k = N$ regenerated samples $\Rightarrow$ SR=NSSR=I-SIR
- **Statistical properties:** Given the previous set of particles $\{X^i_{0:n-1}\}_{i=1}^N$, for all k, $0 \leq k \leq N$, we have:

$$\mathrm{E}(\widehat{\Theta}_t^{\mathrm{NSSR},N,k}) = \mathrm{E}(\widehat{\Theta}_t^{\mathrm{SR},N,k}) = \mathrm{E}(\widehat{\Theta}_t^{\mathrm{I-SIR},N}) = \mathrm{E}(\widehat{\Theta}_t^{\mathrm{SIR},N})$$

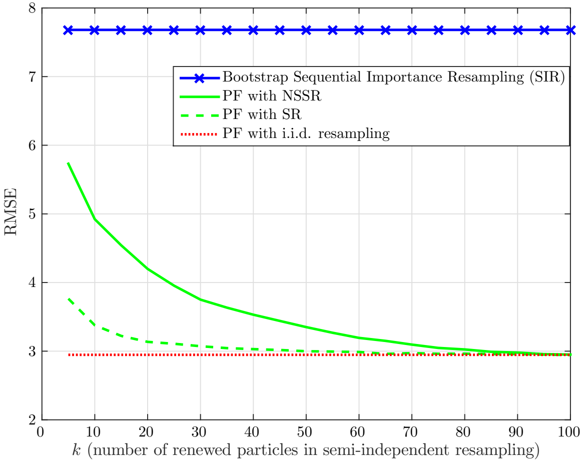
$$\mathrm{var}(\widehat{\Theta}_t^{\mathrm{I-SIR},N}) \leq \mathrm{var}(\widehat{\Theta}_t^{\mathrm{SR},N,k}) \leq \mathrm{var}(\widehat{\Theta}_t^{\mathrm{NSSR},N,k}) \leq \mathrm{var}(\widehat{\Theta}_t^{\mathrm{SIR},N})$$

$$\mathrm{var}(\widehat{\Theta}_t^{\mathrm{SR},N,k}) \leq \mathrm{var}(\widehat{\Theta}_t^{\mathrm{SR},N,k-1})$$

$$\mathrm{var}(\widehat{\Theta}_t^{\mathrm{NSSR},N,k}) \leq \mathrm{var}(\widehat{\Theta}_t^{\mathrm{NSSR},N,k-1})$$

Numerical simulations - Tracking

Target tracking scenario: range-bearing measurements $(\sigma_Q = \sqrt{10},\ \sigma_\rho = 0.1,\ \sigma_\theta = \frac{\pi}{1800})$

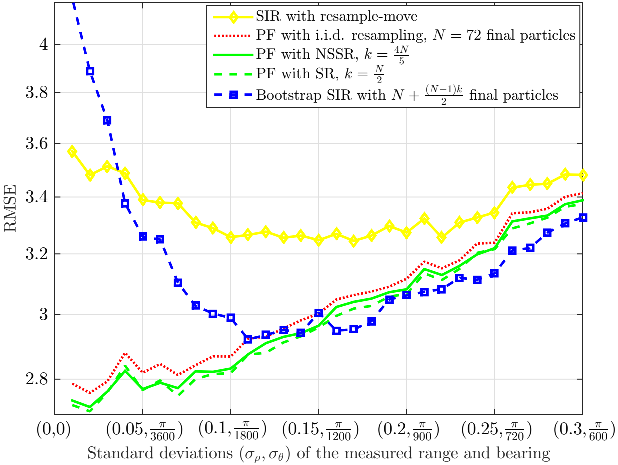


- Improvement when k increases
- note that $\widehat{\Theta}_t^{\mathrm{SR},N,k}$ (resp. $\widehat{\Theta}_t^{\mathrm{NSSR},N}$) has almost the same performance as $\widehat{\Theta}_t^{\mathrm{I-SIR},N}$ when $k \geq N/2$ (resp. $k \geq 4N/5$)

Numerical simulations - Tracking

Target tracking scenario: range-bearing measurements ($\sigma_Q = \sqrt{10}$)

Comparison with other algorithms, each having the same computational cost



- Significant gain for informative models , i.e. small values for $(\sigma_\rho, \sigma_\theta)$

Conclusion and Perspectives

Conclusion

- Revisit of sampling/resampling as a compound IS distribution
  - Proposition of a general technique to draw (semi-)independent samples
  - Theoretical analysis showing the benefit of such an approach
  - At equivalent cost, the proposed approach outperforms the existing techniques in highly informative/high-dimensional models

Conclusion and Perspectives

Conclusion

- Revisit of sampling/resampling as a compound IS distribution
  - Proposition of a general technique to draw (semi-)independent samples
  - Theoretical analysis showing the benefit of such an approach
  - At equivalent cost, the proposed approach outperforms the existing techniques in highly informative/high-dimensional models

Perspectives

- Optimization of k (# regenerated samples): trade-off performance & cost
- Non-uniform selection of samples to regenerate