Sequential Importance Sampling Resampling

Arnaud Doucet
Departments of Statistics & Computer Science
University of British Columbia

Sequential Importance Sampling

- We use a structured IS distribution

$$\begin{aligned} q_n(x_{1:n}) &= q_{n-1}(x_{1:n-1})\, q_n(x_n | x_{1:n-1}) \\ &= q_1(x_1)\, q_2(x_2 | x_1) \cdots q_n(x_n | x_{1:n-1}) \end{aligned}$$

so if $X_{1:n-1}^{(i)} \sim q_{n-1}(x_{1:n-1})$ then we only need to sample $X_n^{(i)} \Big| X_{1:n-1}^{(i)} \sim q_n\left(x_n | X_{1:n-1}^{(i)}\right)$ to obtain $X_{1:n}^{(i)} \sim q_n(x_{1:n})$

Sequential Importance Sampling

- We use a structured IS distribution

$$\begin{aligned} q_n(x_{1:n}) &= q_{n-1}(x_{1:n-1})\, q_n(x_n | x_{1:n-1}) \\ &= q_1(x_1)\, q_2(x_2 | x_1) \cdots q_n(x_n | x_{1:n-1}) \end{aligned}$$

so if $X_{1:n-1}^{(i)} \sim q_{n-1}(x_{1:n-1})$ then we only need to sample $X_n^{(i)} \Big| X_{1:n-1}^{(i)} \sim q_n\left(x_n | X_{1:n-1}^{(i)}\right)$ to obtain $X_{1:n}^{(i)} \sim q_n(x_{1:n})$

- The importance weights are updated according to

$$w_n(x_{1:n}) = \frac{\gamma_n(x_{1:n})}{q_n(x_{1:n})} = w_{n-1}(x_{1:n-1}) \underbrace{\frac{\gamma_n(x_{1:n})}{\gamma_{n-1}(x_{1:n-1})\, q_n(x_n | x_{1:n-1})}}_{\alpha_n(x_{1:n})}$$

Sequential Importance Sampling

- At time $n = 1$, sample $X_1^{(i)} \sim q_1(\cdot)$ and set $w_1\left(X_1^{(i)}\right) = \frac{\gamma_1\left(X_1^{(i)}\right)}{q_1\left(X_1^{(i)}\right)}$.

Sequential Importance Sampling

- At time $n = 1$, sample $X_1^{(i)} \sim q_1(\cdot)$ and set $w_1\left(X_1^{(i)}\right) = \frac{\gamma_1\left(X_1^{(i)}\right)}{q_1\left(X_1^{(i)}\right)}$.
- At time $n \geq 2$

Sequential Importance Sampling

- At time $n = 1$, sample $X_1^{(i)} \sim q_1(\cdot)$ and set $w_1\left(X_1^{(i)}\right) = \frac{\gamma_1\left(X_1^{(i)}\right)}{q_1\left(X_1^{(i)}\right)}$.
- At time $n \geq 2$
  - sample $X_n^{(i)} \sim q_n\left(\cdot \middle| X_{1:n-1}^{(i)}\right)$

Sequential Importance Sampling

- At time $n=1$, sample $X_1^{(i)} \sim q_1(\cdot)$ and set $w_1\left(X_1^{(i)}\right) = \frac{\gamma_1\left(X_1^{(i)}\right)}{q_1\left(X_1^{(i)}\right)}$.
- At time $n \geq 2$
  - sample $X_n^{(i)} \sim q_n\left(\cdot \middle| X_{1:n-1}^{(i)}\right)$
  - compute $w_n\left(X_{1:n}^{(i)}\right) = w_{n-1}\left(X_{1:n-1}^{(i)}\right) \alpha_n\left(X_{1:n}^{(i)}\right)$.

Sequential Importance Sampling

- At time $n=1$, sample $X_1^{(i)} \sim q_1(\cdot)$ and set $w_1\left(X_1^{(i)}\right) = \frac{\gamma_1\left(X_1^{(i)}\right)}{q_1\left(X_1^{(i)}\right)}$.
- At time $n \geq 2$
  - sample $X_n^{(i)} \sim q_n\left(\cdot \middle| X_{1:n-1}^{(i)}\right)$
  - compute $w_n\left(X_{1:n}^{(i)}\right) = w_{n-1}\left(X_{1:n-1}^{(i)}\right) \alpha_n\left(X_{1:n}^{(i)}\right)$.
- It follows that

$$\begin{aligned} \widehat{\pi}_n(dx_{1:n}) &= \sum_{i=1}^{N} W_n^{(i)} \delta_{X_{1:n}^{(i)}}(dx_{1:n}), \\ \widehat{Z}_n &= \frac{1}{N} \sum_{i=1}^{N} w_n\left(X_{1:n}^{(i)}\right). \end{aligned}$$

Sequential Importance Sampling for State-Space Models

- State-space models

$$\text{Hidden Markov process: } X_1 \sim \mu, \;\; X_k | (X_{k-1} = x_{k-1}) \sim f(\cdot | x_{k-1})$$

$$\text{Observation process: } \;\; Y_k | (X_k = x_k) \sim g(\cdot | x_k)$$

Sequential Importance Sampling for State-Space Models

- State-space models

$$\text{Hidden Markov process: } X_1 \sim \mu, \;\; X_k | \left(X_{k-1} = x_{k-1}\right) \sim f\left(\cdot | x_{k-1}\right)$$

$$\text{Observation process: } \;\; Y_k | \left(X_k = x_k\right) \sim g\left(\cdot | x_k\right)$$

- Assume we have received $y_{1:n}$, we are interested in sampling from

$$\pi_n\left(x_{1:n}\right) = p\left(x_{1:n} | y_{1:n}\right) = \frac{p\left(x_{1:n}, y_{1:n}\right)}{p\left(y_{1:n}\right)}$$

and estimating $p\left(y_{1:n}\right)$ where

$$\gamma_n\left(x_{1:n}\right) = p\left(x_{1:n}, y_{1:n}\right) = \mu\left(x_1\right) \prod_{k=2}^{n} f\left(x_k | x_{k-1}\right) \prod_{k=1}^{n} g\left(y_k | x_k\right),$$

$$Z_n = p\left(y_{1:n}\right) = \int \cdots \int \mu\left(x_1\right) \prod_{k=2}^{n} f\left(x_k | x_{k-1}\right) \prod_{k=1}^{n} g\left(y_k | x_k\right) dx_{1:n}.$$

Locally Optimal Importance Distribution

- The optimal IS distribution $q_n(x_n | x_{1:n-1})$ at time n minimizing the variance of $w_n(x_{1:n})$ is given by

$$q_n^{\text{opt}}(x_n | x_{1:n-1}) = \pi_n(x_n | x_{1:n-1})$$

and yields an incremental importance weight of the form

$$\alpha_n(x_{1:n}) = \frac{\gamma_n(x_{1:n-1})}{\gamma_{n-1}(x_{1:n-1})}$$

Locally Optimal Importance Distribution

- The optimal IS distribution $q_n(x_n | x_{1:n-1})$ at time n minimizing the variance of $w_n(x_{1:n})$ is given by

$$q_n^{\text{opt}}(x_n | x_{1:n-1}) = \pi_n(x_n | x_{1:n-1})$$

  and yields an incremental importance weight of the form

$$\alpha_n(x_{1:n}) = \frac{\gamma_n(x_{1:n-1})}{\gamma_{n-1}(x_{1:n-1})}$$

- For state-space models, we have

$$q_n^{\text{opt}}(x_n | x_{1:n-1}) = p(x_n | y_n, x_{n-1}) = \frac{g(y_n | x_n) f(x_n | x_{n-1})}{p(y_n | x_{n-1})},$$

$$\alpha_n(x_{1:n}) = p(y_n | x_{n-1}).$$

Summary

- Sequential Importance Sampling is a special case of Importance Sampling.

Summary

- Sequential Importance Sampling is a special case of Importance Sampling.
- Importance Sampling only works decently for moderate size problems.

Summary

- Sequential Importance Sampling is a special case of Importance Sampling.
- Importance Sampling only works decently for moderate size problems.
- Today, we discuss how to **partially** fix this problem.

Resampling

- *Intuitive KEY idea*: As the time index n increases, the variance of the unnormalized weights $\left\{ w_n \left(X_{1:n}^{(i)} \right) \right\}$ tend to increase and all the mass is concentrated on a few random samples/particles. We propose to reset the approximation by getting rid in a principled way of the particles with low weights $W_n^{(i)}$ (relative to $1/N$) and multiply the particles with high weights $W_n^{(i)}$ (relative to $1/N$).

Resampling

- *Intuitive KEY idea*: As the time index n increases, the variance of the unnormalized weights $\left\{ w_n \left(X_{1:n}^{(i)} \right) \right\}$ tend to increase and all the mass is concentrated on a few random samples/particles. We propose to reset the approximation by getting rid in a principled way of the particles with low weights $W_n^{(i)}$ (relative to $1/N$) and multiply the particles with high weights $W_n^{(i)}$ (relative to $1/N$).
- The main reason is that if a particle at time n has a low weight then typically it will still have a low weight at time $n+1$ (though I can easily give you a counterexample).

Resampling

- *Intuitive KEY idea*: As the time index n increases, the variance of the unnormalized weights $\left\{ w_n \left(X_{1:n}^{(i)} \right) \right\}$ tend to increase and all the mass is concentrated on a few random samples/particles. We propose to reset the approximation by getting rid in a principled way of the particles with low weights $W_n^{(i)}$ (relative to $1/N$) and multiply the particles with high weights $W_n^{(i)}$ (relative to $1/N$).
- The main reason is that if a particle at time n has a low weight then typically it will still have a low weight at time $n+1$ (though I can easily give you a counterexample).
- You want to focus your computational efforts on the "promising" parts of the space.

Multinomial Resampling

- At time n, IS provides the following approximation of $\pi_n\left(x_{1:n}\right)$

$$\widehat{\pi}_n\left(dx_{1:n}\right)=\sum_{i=1}^{N}W_n^{(i)}\delta_{X_{1:n}^{(i)}}\left(dx_{1:n}\right).$$

Multinomial Resampling

- At time n, IS provides the following approximation of $\pi_n(x_{1:n})$

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

- The simplest resampling schemes consists of sampling N times $\widetilde{X}_{1:n}^{(i)} \sim \widehat{\pi}_n(dx_{1:n})$ to build the new approximation

$$\widetilde{\pi}_n(dx_{1:n}) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\widetilde{X}_{1:n}^{(i)}}(dx_{1:n}).$$

Multinomial Resampling

- At time n, IS provides the following approximation of $\pi_n(x_{1:n})$

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

- The simplest resampling schemes consists of sampling N times $\widetilde{X}_{1:n}^{(i)} \sim \widehat{\pi}_n(dx_{1:n})$ to build the new approximation

$$\widetilde{\pi}_n(dx_{1:n}) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\widetilde{X}_{1:n}^{(i)}}(dx_{1:n}).$$

- The new resampled particles $\left\{\widetilde{X}_{1:n}^{(i)}\right\}$ are *approximately* distributed according to $\pi_n(x_{1:n})$ but statistically dependent. This is theoretically more difficult to study.

- Note that we can rewrite

$$\widetilde{\pi}_n\left(dx_{1:n}\right)=\sum_{i=1}^{N}\frac{N_n^{(i)}}{N}\delta_{X_{1:n}^{(i)}}\left(dx_{1:n}\right)$$

where $\left(N_n^{(1)},...,N_n^{(N)}\right)\sim\mathcal{M}\left(N;W_n^{(1)},...,W_n^{(N)}\right)$ thus $\mathbb{E}\left[N_n^{(i)}\right]=NW_n^{(i)},\ var\left[N_n^{(1)}\right]=NW_n^{(i)}\left(1-W_n^{(i)}\right).$

- Note that we can rewrite

$$\widetilde{\pi}_n(dx_{1:n}) = \sum_{i=1}^{N} \frac{N_n^{(i)}}{N} \delta_{X_{1:n}^{(i)}}(dx_{1:n})$$

  where $\left(N_n^{(1)}, ..., N_n^{(N)}\right) \sim \mathcal{M}\left(N; W_n^{(1)}, ..., W_n^{(N)}\right)$ thus $\mathbb{E}\left[N_n^{(i)}\right] = NW_n^{(i)}$, $var\left[N_n^{(1)}\right] = NW_n^{(i)}\left(1 - W_n^{(i)}\right)$.
- It follows that the resampling step is an unbiased operation

$$\mathbb{E}\left[\widetilde{\pi}_n(dx_{1:n}) \middle| \widehat{\pi}_n(dx_{1:n})\right] = \widehat{\pi}_n(dx_{1:n})$$

  but clearly it introduces some errors "locally" in time. That is for any test function, we have

$$var_{\widetilde{\pi}_n}[\varphi(X_{1:n})] \geq var_{\widehat{\pi}_n}[\varphi(X_{1:n})]$$

- Note that we can rewrite

$$\widetilde{\pi}_n\left(dx_{1:n}\right)=\sum_{i=1}^{N}\frac{N_n^{(i)}}{N}\delta_{X_{1:n}^{(i)}}\left(dx_{1:n}\right)$$

where $\left(N_n^{(1)},...,N_n^{(N)}\right)\sim\mathcal{M}\left(N;W_n^{(1)},...,W_n^{(N)}\right)$ thus $\mathbb{E}\left[N_n^{(i)}\right]=NW_n^{(i)}$, $var\left[N_n^{(1)}\right]=NW_n^{(i)}\left(1-W_n^{(i)}\right)$.

- It follows that the resampling step is an unbiased operation

$$\mathbb{E}\left[\left.\widetilde{\pi}_n\left(dx_{1:n}\right)\right|\widehat{\pi}_n\left(dx_{1:n}\right)\right]=\widehat{\pi}_n\left(dx_{1:n}\right)$$

but clearly it introduces some errors "locally" in time. That is for any test function, we have

$$var_{\widetilde{\pi}_n}\left[\varphi\left(X_{1:n}\right)\right]\geq var_{\widehat{\pi}_n}\left[\varphi\left(X_{1:n}\right)\right]$$

- Resampling is beneficial for future time steps (sometimes).

Stratified Resampling

- Better resampling steps can be designed such that $\mathbb{E}\left[N_n^{(i)}\right] = NW_n^{(i)}$ but $\mathbb{V}\left[N_n^{(i)}\right] < NW_n^{(i)}\left(1 - W_n^{(i)}\right)$.

Stratified Resampling

- Better resampling steps can be designed such that $\mathbb{E}\left[N_n^{(i)}\right] = NW_n^{(i)}$ but $\mathbb{V}\left[N_n^{(i)}\right] < NW_n^{(i)}\left(1 - W_n^{(i)}\right)$.
- A popular alternative to multinomial resampling consists of selecting

$$U_1 \sim \mathcal{U}\left[0, \frac{1}{N}\right]$$

and for $i = 2, ..., N$

$$U_i = U_1 + \frac{i-1}{N} = U_{i-1} + \frac{1}{N}.$$

Stratified Resampling

- Better resampling steps can be designed such that $\mathbb{E}\left[N_n^{(i)}\right] = NW_n^{(i)}$ but $\mathbb{V}\left[N_n^{(i)}\right] < NW_n^{(i)}\left(1 - W_n^{(i)}\right)$.
- A popular alternative to multinomial resampling consists of selecting

$$U_1 \sim \mathcal{U}\left[0, \frac{1}{N}\right]$$

and for $i = 2, ..., N$

$$U_i = U_1 + \frac{i-1}{N} = U_{i-1} + \frac{1}{N}.$$

- Then we set

$$N_n^{(i)} = \#\left\{U_j : \sum_{m=1}^{i-1} W_n^{(m)} \leq U_j < \sum_{m=1}^{i} W_n^{(m)}\right\}$$

where $\sum_{m=1}^{0} = 0$.

Stratified Resampling

- Better resampling steps can be designed such that $\mathbb{E}\left[N_n^{(i)}\right] = NW_n^{(i)}$ but $\mathbb{V}\left[N_n^{(i)}\right] < NW_n^{(i)}\left(1 - W_n^{(i)}\right)$.
- A popular alternative to multinomial resampling consists of selecting

$$U_1 \sim \mathcal{U}\left[0, \frac{1}{N}\right]$$

and for $i = 2, ..., N$

$$U_i = U_1 + \frac{i-1}{N} = U_{i-1} + \frac{1}{N}.$$

- Then we set

$$N_n^{(i)} = \#\left\{U_j : \sum_{m=1}^{i-1} W_n^{(m)} \leq U_j < \sum_{m=1}^{i} W_n^{(m)}\right\}$$

where $\sum_{m=1}^{0} = 0$.
- It is trivial to check that $\mathbb{E}\left[N_n^{(i)}\right] = NW_n^{(i)}$.

An alternative approach to resampling

- Assume

$$\alpha_n\left(x_{1:n}\right) \leq 1 \text{ over } E_n \text{ (rescale if necessary)}$$

An alternative approach to resampling

- Assume

$$\alpha_n(x_{1:n}) \leq 1 \text{ over } E_n \text{ (rescale if necessary)}$$

- We have

$$\begin{aligned}
\pi_n(x_{1:n}) &= \frac{\alpha_n(x_{1:n})\, q_n(x_n | x_{1:n-1})\, \pi_{n-1}(x_{1:n-1})}{\int \alpha_n(x_{1:n})\, q_n(dx_n | x_{1:n-1})\, \pi_{n-1}(dx_{1:n-1})} \\
&= \alpha_n(x_{1:n})\, q_n(x_n | x_{1:n-1})\, \pi_{n-1}(x_{1:n-1}) \\
&\quad + \left(1 - \int \alpha_n(x_{1:n})\, q_n(dx_n | x_{1:n-1})\, \pi_{n-1}(dx_{1:n-1})\right) \\
&\quad \times \frac{\alpha_n(x_{1:n})\, q_n(x_n | x_{1:n-1})\, \pi_{n-1}(x_{1:n-1})}{\int \alpha_n(x_{1:n})\, q_n(dx_n | x_{1:n-1})\, \pi_{n-1}(dx_{1:n-1})}
\end{aligned}$$

An alternative approach to resampling

- Assume
$$\alpha_n(x_{1:n}) \leq 1 \text{ over } E_n \text{ (rescale if necessary)}$$
- We have
$$\begin{aligned}
\pi_n(x_{1:n}) &= \frac{\alpha_n(x_{1:n})\, q_n(x_n|\, x_{1:n-1})\, \pi_{n-1}(x_{1:n-1})}{\int \alpha_n(x_{1:n})\, q_n(dx_n|\, x_{1:n-1})\, \pi_{n-1}(dx_{1:n-1})} \\
&= \alpha_n(x_{1:n})\, q_n(x_n|\, x_{1:n-1})\, \pi_{n-1}(x_{1:n-1}) \\
&\quad + \left(1 - \int \alpha_n(x_{1:n})\, q_n(dx_n|\, x_{1:n-1})\, \pi_{n-1}(dx_{1:n-1})\right) \\
&\quad \times \frac{\alpha_n(x_{1:n})\, q_n(x_n|\, x_{1:n-1})\, \pi_{n-1}(x_{1:n-1})}{\int \alpha_n(x_{1:n})\, q_n(dx_n|\, x_{1:n-1})\, \pi_{n-1}(dx_{1:n-1})}
\end{aligned}$$
- Looks like measure-valued Metropolis-Hastings algorithm.

Probabilistic interpretation

- We have

$$
\begin{aligned}
\pi_n(x_{1:n}) \quad = \quad & \underbrace{\alpha_n(x_{1:n})}_{\text{accept with proba } w_n} \quad \underbrace{q_n(x_n|\, x_{1:n-1})\, \pi_{n-1}(x_{1:n-1})}_{\text{trial distribution}} + \\
& \underbrace{\left(1 - \int \alpha_n(x_{1:n})\, q_n(dx_n|\, x_{1:n-1})\, \pi_{n-1}(dx_{1:n-1})\right)}_{\text{rejection probability}} \pi_n(x_{1:n})
\end{aligned}
$$

Probabilistic interpretation

- We have

$$\pi_n(x_{1:n}) = \underbrace{\alpha_n(x_{1:n})}_{\text{accept with proba } w_n} \underbrace{q_n(x_n | x_{1:n-1})\, \pi_{n-1}(x_{1:n-1})}_{\text{trial distribution}} + \underbrace{\left(1 - \int \alpha_n(x_{1:n})\, q_n(dx_n | x_{1:n-1})\, \pi_{n-1}(dx_{1:n-1})\right)}_{\text{rejection probability}} \pi_n(x_{1:n})$$

- Say $X_{1:n-1}^{(i)} \sim \pi_{n-1}$ and sample $X_n^{(i)} \sim q_n\left(\cdot | X_{1:n-1}^{(i)}\right)$.

Probabilistic interpretation

- We have

$$\pi_n(x_{1:n}) = \underbrace{\alpha_n(x_{1:n})}_{\text{accept with proba } w_n} \underbrace{q_n(x_n | x_{1:n-1})\,\pi_{n-1}(x_{1:n-1})}_{\text{trial distribution}} + \underbrace{\left(1 - \int \alpha_n(x_{1:n})\, q_n(dx_n | x_{1:n-1})\,\pi_{n-1}(dx_{1:n-1})\right)}_{\text{rejection probability}} \pi_n(x_{1:n})$$

- Say $X_{1:n-1}^{(i)} \sim \pi_{n-1}$ and sample $X_n^{(i)} \sim q_n\left(\cdot | X_{1:n-1}^{(i)}\right)$.
- With probability $\alpha_n\left(X_n^{(i)}\right)$, set $\widetilde{X}_{1:n}^{(i)} = X_{1:n}^{(i)}$ otherwise $\widetilde{X}_{1:n}^{(i)} \sim \sum_{i=1}^{N} W_n^{(i)} \delta_{X_{1:n}^{(i)}}(dx_{1:n})$.

Probabilistic interpretation

- We have

$$\pi_n(x_{1:n}) = \underbrace{\alpha_n(x_{1:n})}_{\text{accept with proba } w_n} \underbrace{q_n(x_n | x_{1:n-1})\,\pi_{n-1}(x_{1:n-1})}_{\text{trial distribution}} + \underbrace{\left(1 - \int \alpha_n(x_{1:n})\, q_n(dx_n | x_{1:n-1})\,\pi_{n-1}(dx_{1:n-1})\right)}_{\text{rejection probability}} \pi_n(x_{1:n})$$

- Say $X_{1:n-1}^{(i)} \sim \pi_{n-1}$ and sample $X_n^{(i)} \sim q_n\left(\cdot | X_{1:n-1}^{(i)}\right)$.
- With probability $\alpha_n\left(X_n^{(i)}\right)$, set $\widetilde{X}_{1:n}^{(i)} = X_{1:n}^{(i)}$ otherwise $\widetilde{X}_{1:n}^{(i)} \sim \sum_{i=1}^{N} W_n^{(i)} \delta_{X_{1:n}^{(i)}}(dx_{1:n})$.
- **Remark**: Allows to decrease variance if $\alpha_n(x_{1:n})$ "flat" over E_n; e.g. filtering with large observation noise.

Degeneracy Measures

- Resampling at each time step is harmful. We should resample only when necessary.

Degeneracy Measures

- Resampling at each time step is harmful. We should resample only when necessary.
- To measure the variation of the weights, we can use the Effective Sample Size (ESS) or the coefficient of variation CV

$$ESS = \left(\sum_{i=1}^{N} \left(W_n^{(i)} \right)^2 \right)^{-1}, \quad CV = \left(\frac{1}{N} \sum_{i=1}^{N} \left(N W_n^{(i)} - 1 \right)^2 \right)^{1/2}$$

Degeneracy Measures

- Resampling at each time step is harmful. We should resample only when necessary.
- To measure the variation of the weights, we can use the Effective Sample Size (ESS) or the coefficient of variation CV

$$ESS = \left(\sum_{i=1}^{N} \left(W_n^{(i)} \right)^2 \right)^{-1}, \quad CV = \left(\frac{1}{N} \sum_{i=1}^{N} \left(N W_n^{(i)} - 1 \right)^2 \right)^{1/2}$$

- We have $ESS = N$ and $CV = 0$ if $W_n^{(i)} = 1/N$ for any i.

Degeneracy Measures

- Resampling at each time step is harmful. We should resample only when necessary.
- To measure the variation of the weights, we can use the Effective Sample Size (ESS) or the coefficient of variation CV

$$ESS = \left(\sum_{i=1}^{N} \left(W_n^{(i)} \right)^2 \right)^{-1}, \quad CV = \left(\frac{1}{N} \sum_{i=1}^{N} \left(N W_n^{(i)} - 1 \right)^2 \right)^{1/2}$$

- We have $ESS = N$ and $CV = 0$ if $W_n^{(i)} = 1/N$ for any i.
- We have $ESS = 1$ and $CV = \sqrt{N-1}$ if $W_n^{(i)} = 1$ and $W_n^{(j)} = 1$ for $j \neq i$.

- We can also use the entropy

$$Ent = -\sum_{i=1}^{N} W_n^{(i)} \log_2 \left(W_n^{(i)} \right)$$

- We can also use the entropy

$$Ent = -\sum_{i=1}^{N} W_n^{(i)} \log_2 \left(W_n^{(i)} \right)$$

- We have $Ent = \log_2 (N)$ if $W_n^{(i)} = 1/N$ for any i. We have $Ent = 0$ if $W_n^{(i)} = 1$ and $W_n^{(j)} = 1$ for $j \neq i$.

- We can also use the entropy

$$Ent = -\sum_{i=1}^{N} W_n^{(i)} \log_2 \left(W_n^{(i)} \right)$$

- We have $Ent = \log_2 (N)$ if $W_n^{(i)} = 1/N$ for any i. We have $Ent = 0$ if $W_n^{(i)} = 1$ and $W_n^{(j)} = 1$ for $j \neq i$.
- **Dynamic Resampling**: If the variation of the weights as measured by ESS, CV or Ent is too high, then resample the particles.

Generic Sequential Monte Carlo Scheme

- At time $n = 1$, sample $X_1^{(i)} \sim q_1(\cdot)$ and set $w_1\left(X_1^{(i)}\right) = \frac{\gamma_1\left(X_1^{(i)}\right)}{q_1\left(X_1^{(i)}\right)}$.

Generic Sequential Monte Carlo Scheme

- At time $n = 1$, sample $X_1^{(i)} \sim q_1(\cdot)$ and set $w_1\left(X_1^{(i)}\right) = \frac{\gamma_1\left(X_1^{(i)}\right)}{q_1\left(X_1^{(i)}\right)}$.
- Resample $\left\{X_1^{(i)}, W_1^{(i)}\right\}$ to obtain new particles also denoted $\left\{X_1^{(i)}\right\}$

Generic Sequential Monte Carlo Scheme

- At time $n = 1$, sample $X_1^{(i)} \sim q_1(\cdot)$ and set $w_1\left(X_1^{(i)}\right) = \frac{\gamma_1\left(X_1^{(i)}\right)}{q_1\left(X_1^{(i)}\right)}$.
- Resample $\left\{X_1^{(i)}, W_1^{(i)}\right\}$ to obtain new particles also denoted $\left\{X_1^{(i)}\right\}$
- At time $n \geq 2$

Generic Sequential Monte Carlo Scheme

- At time $n = 1$, sample $X_1^{(i)} \sim q_1(\cdot)$ and set $w_1\left(X_1^{(i)}\right) = \frac{\gamma_1\left(X_1^{(i)}\right)}{q_1\left(X_1^{(i)}\right)}$.
- Resample $\left\{X_1^{(i)}, W_1^{(i)}\right\}$ to obtain new particles also denoted $\left\{X_1^{(i)}\right\}$
- At time $n \geq 2$
  - sample $X_n^{(i)} \sim q_n\left(\cdot\middle|\, X_{1:n-1}^{(i)}\right)$

Generic Sequential Monte Carlo Scheme

- At time $n=1$, sample $X_1^{(i)} \sim q_1(\cdot)$ and set $w_1\left(X_1^{(i)}\right) = \frac{\gamma_1\left(X_1^{(i)}\right)}{q_1\left(X_1^{(i)}\right)}$.
- Resample $\left\{X_1^{(i)}, W_1^{(i)}\right\}$ to obtain new particles also denoted $\left\{X_1^{(i)}\right\}$
- At time $n \geq 2$
  - sample $X_n^{(i)} \sim q_n\left(\cdot \middle| X_{1:n-1}^{(i)}\right)$
  - compute $w_n\left(X_{1:n}^{(i)}\right) = \alpha_n\left(X_{1:n}^{(i)}\right)$.

Generic Sequential Monte Carlo Scheme

- At time $n=1$, sample $X_1^{(i)} \sim q_1(\cdot)$ and set $w_1\left(X_1^{(i)}\right) = \frac{\gamma_1\left(X_1^{(i)}\right)}{q_1\left(X_1^{(i)}\right)}$.
- Resample $\left\{X_1^{(i)}, W_1^{(i)}\right\}$ to obtain new particles also denoted $\left\{X_1^{(i)}\right\}$
- At time $n \geq 2$
  - sample $X_n^{(i)} \sim q_n\left(\cdot \middle| X_{1:n-1}^{(i)}\right)$
  - compute $w_n\left(X_{1:n}^{(i)}\right) = \alpha_n\left(X_{1:n}^{(i)}\right)$.
- Resample $\left\{X_{1:n}^{(i)}, W_n^{(i)}\right\}$ to obtain new particles also denoted $\left\{X_{1:n}^{(i)}\right\}$

- At any time n, we have two approximation of $\pi_n(x_{1:n})$

$$\widehat{\pi}_n(dx_{1:n}) = \sum_{i=1}^{N} W_n^{(i)} \delta_{X_{1:n}^{(i)}}(dx_{1:n}) \text{ (before resampling)}$$

$$\widetilde{\pi}_n(dx_{1:n}) = \frac{1}{N} \sum_{i=1}^{N} \delta_{X_{1:n}^{(i)}}(dx_{1:n}) \text{ (after resampling).}$$

- At any time n, we have two approximation of $\pi_n(x_{1:n})$

$$\begin{aligned}\widehat{\pi}_n(dx_{1:n}) &= \sum_{i=1}^{N} W_n^{(i)} \delta_{X_{1:n}^{(i)}}(dx_{1:n}) \text{ (before resampling)} \\ \widetilde{\pi}_n(dx_{1:n}) &= \frac{1}{N}\sum_{i=1}^{N} \delta_{X_{1:n}^{(i)}}(dx_{1:n}) \text{ (after resampling).}\end{aligned}$$

- We also have

$$\frac{\widehat{Z_n}}{Z_{n-1}} = \frac{1}{N}\sum_{i=1}^{N} w_n\left(X_{1:n}^{(i)}\right).$$

Sequential Monte Carlo for Hidden Markov Models

- At time $n = 1$, sample $X_1^{(i)} \sim q_1(\cdot)$ and set $w_1\left(X_1^{(i)}\right) = \frac{\mu\left(X_1^{(i)}\right) g\left(y_1 | X_1^{(i)}\right)}{q\left(X_1^{(i)} \middle| y_1\right)}$.

Sequential Monte Carlo for Hidden Markov Models

- At time $n = 1$, sample $X_1^{(i)} \sim q_1(\cdot)$ and set $w_1\left(X_1^{(i)}\right) = \frac{\mu\left(X_1^{(i)}\right) g\left(y_1 | X_1^{(i)}\right)}{q\left(X_1^{(i)} \middle| y_1\right)}$.
- Resample $\left\{X_1^{(i)}, W_1^{(i)}\right\}$ to obtain new particles also denoted $\left\{X_1^{(i)}\right\}$

Sequential Monte Carlo for Hidden Markov Models

- At time $n = 1$, sample $X_1^{(i)} \sim q_1(\cdot)$ and set $w_1\left(X_1^{(i)}\right) = \frac{\mu\left(X_1^{(i)}\right) g\left(y_1 | X_1^{(i)}\right)}{q\left(X_1^{(i)} \middle| y_1\right)}$.
- Resample $\left\{X_1^{(i)}, W_1^{(i)}\right\}$ to obtain new particles also denoted $\left\{X_1^{(i)}\right\}$
- At time $n \geq 2$

Sequential Monte Carlo for Hidden Markov Models

- At time $n = 1$, sample $X_1^{(i)} \sim q_1(\cdot)$ and set
  $w_1\left(X_1^{(i)}\right) = \frac{\mu\left(X_1^{(i)}\right) g\left(y_1 | X_1^{(i)}\right)}{q\left(X_1^{(i)} \middle| y_1\right)}$.
- Resample $\left\{X_1^{(i)}, W_1^{(i)}\right\}$ to obtain new particles also denoted $\left\{X_1^{(i)}\right\}$
- At time $n \geq 2$
  - sample $X_n^{(i)} \sim q\left(\cdot | \, y_n, X_{n-1}^{(i)}\right)$

Sequential Monte Carlo for Hidden Markov Models

- At time $n = 1$, sample $X_1^{(i)} \sim q_1(\cdot)$ and set $w_1\left(X_1^{(i)}\right) = \frac{\mu\left(X_1^{(i)}\right) g\left(y_1 | X_1^{(i)}\right)}{q\left(X_1^{(i)} \middle| y_1\right)}$.
- Resample $\left\{X_1^{(i)}, W_1^{(i)}\right\}$ to obtain new particles also denoted $\left\{X_1^{(i)}\right\}$
- At time $n \geq 2$
  - sample $X_n^{(i)} \sim q\left(\cdot | y_n, X_{n-1}^{(i)}\right)$
  - compute $w_n\left(X_{1:n}^{(i)}\right) = \frac{f\left(X_n^{(i)} \middle| X_{n-1}^{(i)}\right) g\left(y_n | X_n^{(i)}\right)}{q\left(X_n^{(i)} \middle| y_n, X_{n-1}^{(i)}\right)}$.

Sequential Monte Carlo for Hidden Markov Models

- At time $n = 1$, sample $X_1^{(i)} \sim q_1(\cdot)$ and set $w_1\left(X_1^{(i)}\right) = \frac{\mu\left(X_1^{(i)}\right) g\left(y_1 | X_1^{(i)}\right)}{q\left(X_1^{(i)} \middle| y_1\right)}$.
- Resample $\left\{X_1^{(i)}, W_1^{(i)}\right\}$ to obtain new particles also denoted $\left\{X_1^{(i)}\right\}$
- At time $n \geq 2$
  - sample $X_n^{(i)} \sim q\left(\cdot | \, y_n, X_{n-1}^{(i)}\right)$
  - compute $w_n\left(X_{1:n}^{(i)}\right) = \frac{f\left(X_n^{(i)} \middle| X_{n-1}^{(i)}\right) g\left(y_n | X_n^{(i)}\right)}{q\left(X_n^{(i)} \middle| y_n, X_{n-1}^{(i)}\right)}$.
- Resample $\left\{X_{1:n}^{(i)}, W_n^{(i)}\right\}$ to obtain new particles also denoted $\left\{X_{1:n}^{(i)}\right\}$

- **Example**: Linear Gaussian model

$$\begin{aligned} X_1 &\sim \mathcal{N}(0,1), \;\; X_n = \alpha X_{n-1} + \sigma_v V_n, \\ Y_n &= X_n + \sigma_w W_n \end{aligned}$$

where $V_n \sim \mathcal{N}(0,1)$ and $W_n \sim \mathcal{N}(0,1)$.

- **Example**: Linear Gaussian model

$$\begin{aligned} X_1 &\sim \mathcal{N}(0,1), \;\; X_n = \alpha X_{n-1} + \sigma_v V_n, \\ Y_n &= X_n + \sigma_w W_n \end{aligned}$$

  where $V_n \sim \mathcal{N}(0,1)$ and $W_n \sim \mathcal{N}(0,1)$.

- We know that $p(x_{1:n} | y_{1:n})$ is Gaussian and its parameters can be computed using Kalman techniques. In particular $p(x_n | y_{1:n})$ is also a Gaussian which can be computed using the Kalman filter.

- **Example**: Linear Gaussian model

$$
\begin{aligned}
X_1 &\sim \mathcal{N}(0,1), \; X_n = \alpha X_{n-1} + \sigma_v V_n, \\
Y_n &= X_n + \sigma_w W_n
\end{aligned}
$$

  where $V_n \sim \mathcal{N}(0,1)$ and $W_n \sim \mathcal{N}(0,1)$.

- We know that $p(x_{1:n} | y_{1:n})$ is Gaussian and its parameters can be computed using Kalman techniques. In particular $p(x_n | y_{1:n})$ is also a Gaussian which can be computed using the Kalman filter.
- We apply the SMC method with $q(x_n | y_n, x_{n-1}) = f(x_n | x_{n-1}) = \mathcal{N}(x_n; \alpha x_{n-1}, \sigma_v^2)$.

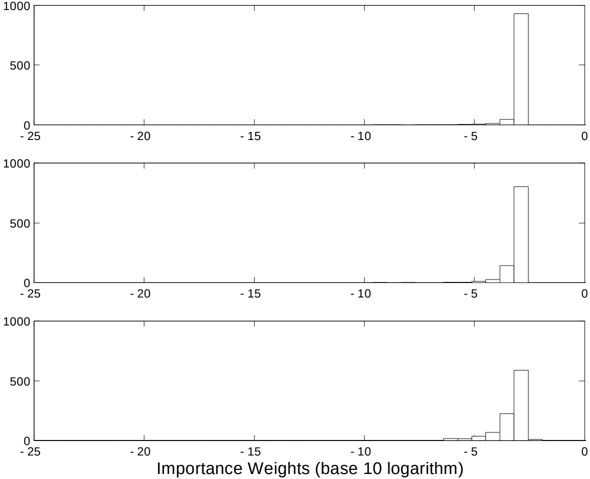

Figure: Histograms of the base 10 logarithm of $W_n^{(i)}$ for $n = 1$ (top), $n = 50$ (middle) and $n = 100$ (bottom).

- By itself this graph does not mean that the procedure is efficient!

- This SMC strategy performs remarkably well in terms of estimation of the marginals $p(x_k | y_{1:k})$. This is what is only necessary in many applications thankfully.

- This SMC strategy performs remarkably well in terms of estimation of the marginals $p(x_k | y_{1:k})$. This is what is only necessary in many applications thankfully.
- However, the joint distribution $p(x_{1:k} | y_{1:k})$ is poorly estimated when k is large; i.e. we have in the previous example

$$\widehat{p}(x_{1:11} | y_{1:24}) = \delta_{X_{1:11}}(x_{1:11}).$$

- This SMC strategy performs remarkably well in terms of estimation of the marginals $p\left(x_k \middle| y_{1:k}\right)$. This is what is only necessary in many applications thankfully.
- However, the joint distribution $p\left(x_{1:k} \middle| y_{1:k}\right)$ is poorly estimated when k is large; i.e. we have in the previous example

$$\widehat{p}\left(x_{1:11} \middle| y_{1:24}\right) = \delta_{X_{1:11}}\left(x_{1:11}\right).$$

- The same conclusion holds for most sequences of distributions $\pi_k\left(x_{1:k}\right)$.

- This SMC strategy performs remarkably well in terms of estimation of the marginals $p(x_k | y_{1:k})$. This is what is only necessary in many applications thankfully.
- However, the joint distribution $p(x_{1:k} | y_{1:k})$ is poorly estimated when k is large; i.e. we have in the previous example

$$\widehat{p}(x_{1:11} | y_{1:24}) = \delta_{X_{1:11}}(x_{1:11}).$$

- The same conclusion holds for most sequences of distributions $\pi_k(x_{1:k})$.
- **Resampling only solves partially our problems**.

Another Illustration of the Degeneracy Phenomenon

- For the linear Gaussian state-space model described before, we can compute in closed form

$$S_n = \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}\left[X_k^2 \middle| Y_{1:n} \right]$$

using the Kalman techniques.

Another Illustration of the Degeneracy Phenomenon

- For the linear Gaussian state-space model described before, we can compute in closed form

$$S_n = \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}\left[X_k^2 \middle| Y_{1:n} \right]$$

using the Kalman techniques.

- We compute the SMC estimate of this quantity given by

$$\widehat{S}_n = \frac{1}{n} \sum_{k=1}^{n} \sum_{i=1}^{N} W_n^{(i)} \left(X_k^{(i)} \right)^2$$

Another Illustration of the Degeneracy Phenomenon

- For the linear Gaussian state-space model described before, we can compute in closed form

$$S_n = \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}\left[X_k^2 \middle| Y_{1:n} \right]$$

  using the Kalman techniques.

- We compute the SMC estimate of this quantity given by

$$\widehat{S}_n = \frac{1}{n} \sum_{k=1}^{n} \sum_{i=1}^{N} W_n^{(i)} \left(X_k^{(i)} \right)^2$$

- This estimate can be updated sequentially using our SMC approximation.

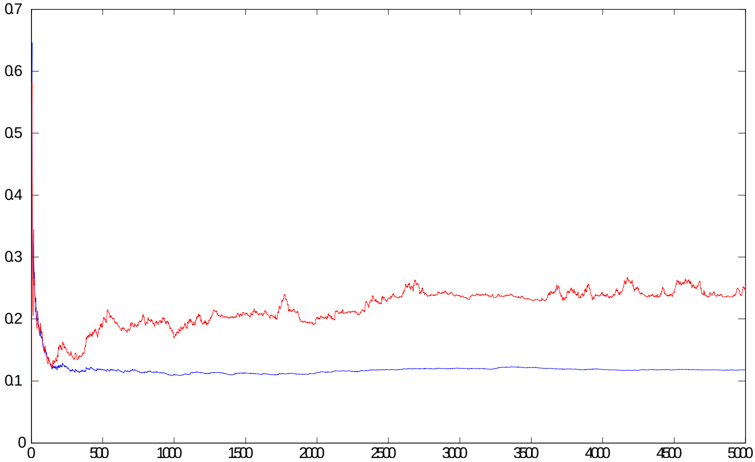

Figure: Sufficient statistics computed exactly through the Kalman smoother (blue) and the SMC method (red).

Some Convergence Results for SMC

- We will discuss convergence results for SMC later; see (Del Moral, 2004).

Some Convergence Results for SMC

- We will discuss convergence results for SMC later; see (Del Moral, 2004).
- In particular we have for any bounded function φ and any $p > 1$

$$\mathbb{E}\left[\left|\int \varphi_n\left(x_{1:n}\right)\left(\widehat{\pi}_n\left(dx_{1:n}\right)-\pi_n\left(dx_{1:n}\right)\right)\right|^p\right]^{1/p} \leq \frac{C_n \left\|\varphi\right\|_\infty}{N}.$$

Some Convergence Results for SMC

- We will discuss convergence results for SMC later; see (Del Moral, 2004).
- In particular we have for any bounded function φ and any $p > 1$

$$\mathbb{E}\left[\left|\int \varphi_n\left(x_{1:n}\right)\left(\widehat{\pi}_n\left(dx_{1:n}\right)-\pi_n\left(dx_{1:n}\right)\right)\right|^p\right]^{1/p} \leq \frac{C_n \left\|\varphi\right\|_\infty}{N}.$$

- It looks like a nice result but it is rather useless as C_n increases polynomially/exponentially with time.

Some Convergence Results for SMC

- We will discuss convergence results for SMC later; see (Del Moral, 2004).
- In particular we have for any bounded function φ and any $p > 1$

$$\mathbb{E}\left[\left|\int \varphi_n\left(x_{1:n}\right)\left(\widehat{\pi}_n\left(dx_{1:n}\right) - \pi_n\left(dx_{1:n}\right)\right)\right|^p\right]^{1/p} \leq \frac{C_n \left\|\varphi\right\|_\infty}{N}.$$

- It looks like a nice result but it is rather useless as C_n increases polynomially/exponentially with time.
- To achieve a fixed precision, this would require to use a time-increasing number of particles N.

- **You cannot hope to estimate with a fixed precision a target distribution of increasing dimension.**

- **You cannot hope to estimate with a fixed precision a target distribution of increasing dimension.**
- At best, you can expect results of the following form

$$\mathbb{E}\left[\left|\int \varphi\left(x_{n-L+1:n}\right)\left(\widehat{\pi}_n\left(dx_{n-L+1:n}\right)-\pi_n\left(dx_{n-L+1:n}\right)\right)\right|^p\right]^{1/p} \leq \frac{M_L \left\|\varphi\right\|_\infty}{N}$$

if the model has nice forgetting/mixing properties, i.e.

$$\int \left|\pi_n\left(x_n \middle| x_1\right)-\pi_n\left(x_n \middle| x_1'\right)\right| dx_n \leq 2\lambda^{n-1}$$

with $0 \leq \lambda < 1$.

- **You cannot hope to estimate with a fixed precision a target distribution of increasing dimension.**
- At best, you can expect results of the following form

$$\mathbb{E}\left[\left|\int \varphi\left(x_{n-L+1:n}\right)\left(\widehat{\pi}_n\left(dx_{n-L+1:n}\right)-\pi_n\left(dx_{n-L+1:n}\right)\right)\right|^p\right]^{1/p} \leq \frac{M_L\left\|\varphi\right\|_{\infty}}{N}$$

if the model has nice forgetting/mixing properties, i.e.

$$\int\left|\pi_n\left(x_n\right|x_1\right)-\pi_n\left(x_n\right|x_1'\right)\left|dx_n \leq 2\lambda^{n-1}$$

with $0 \leq \lambda < 1$.

- In the HMM case, it means that

$$\int\left|p\left(x_n\right|y_{1:n},x_1\right)-p\left(x_n\right|y_{1:n},x_1'\right)\left|dx_n \leq \lambda^{n-1}$$

Central Limit Theorems

- For SIS we have

$$\sqrt{N}\left(\mathbb{E}_{\widehat{\pi}_n}\left(\varphi_n\left(X_{1:n}\right)\right)-\mathbb{E}_{\pi_n}\left(\varphi_n\left(X_{1:n}\right)\right)\right)\Rightarrow\mathcal{N}\left(0,\sigma_{IS}^2\left(\varphi_n\right)\right)$$

where

$$\sigma_{IS}^2\left(\varphi_n\right)=\int\frac{\pi_n^2\left(x_{1:n}\right)}{q_n\left(x_{1:n}\right)}\left(\left(\varphi_n\left(x_{1:n}\right)\right)-\mathbb{E}_{\pi_n}\left(\varphi\left(x_{1:n}\right)\right)\right)^2dx_{1:n}$$

Central Limit Theorems

- For SIS we have

$$\sqrt{N}\left(\mathbb{E}_{\widehat{\pi}_n}\left(\varphi_n\left(X_{1:n}\right)\right)-\mathbb{E}_{\pi_n}\left(\varphi_n\left(X_{1:n}\right)\right)\right)\Rightarrow\mathcal{N}\left(0,\sigma_{IS}^2\left(\varphi_n\right)\right)$$

where

$$\sigma_{IS}^2\left(\varphi_n\right)=\int\frac{\pi_n^2\left(x_{1:n}\right)}{q_n\left(x_{1:n}\right)}\left(\left(\varphi_n\left(x_{1:n}\right)\right)-\mathbb{E}_{\pi_n}\left(\varphi\left(x_{1:n}\right)\right)\right)^2dx_{1:n}$$

- We also have

$$\sqrt{N}\left(\widehat{Z}_n-Z_n\right)\Rightarrow\mathcal{N}\left(0,\sigma_{IS}^2\right)$$

where

$$\sigma_{IS}^2=\int\frac{\pi_n^2\left(x_{1:n}\right)}{q_n\left(x_{1:n}\right)}dx_{1:n}-1$$

- For SMC, we have

$$\sigma^2_{SMC}(\varphi_n) = \int \frac{\pi_n^2(x_1)}{q_1(x_1)} \left(\int \varphi_n(x_{2:n})\, \pi_n(x_{2:n}|\, x_1)\, dx_{2:n} - \mathbb{E}_{\pi_n}(\varphi_n(X_{1:n}))\right)^2 dx_1$$
$$+ \sum_{k=2}^{n-1} \int \frac{\pi_n(x_{1:k})^2}{\pi_{k-1}(x_{1:k-1})\, q_k(x_k|x_{k-1})}$$
$$\times \left(\int \varphi_n(x_{1:n})\, \pi_n(x_{k+1:n}|\, x_k)\, dx_{k+1:n} - \mathbb{E}_{\pi_n}(\varphi_n(X_{1:n}))\right)^2 dx_{1:k}$$
$$+ \int \frac{\pi_n(x_{1:n})^2}{\pi_{n-1}(x_{1:n-1})\, q_n(x_n|x_{n-1})} \left(\varphi_n(x_{1:n}) - \mathbb{E}_{\pi_n}(\varphi_n(X_{1:n}))\right)^2 dx_{1:n}.$$

and

$$\sigma^2_{SMC} = \int \frac{\pi_n^2(x_1)}{q_1(x_1)} dx_1 + \sum_{k=2}^{n} \int \frac{\pi_n(x_{1:k})^2}{\pi_{k-1}(x_{1:k-1})\, q_k(x_k|\, x_{k-1})} dx_{1:k} - n$$

Back to our toy example

- Consider the case where the target is defined on $\mathbb{R}^n$ and

$$\pi\left(x_{1:n}\right) = \prod_{n=1}^{n} \mathcal{N}\left(x_k; 0, 1\right),$$

$$\gamma\left(x_{1:n}\right) = \prod_{k=1}^{n} \exp\left(-\frac{x_k^2}{2}\right), \; Z = \left(2\pi\right)^{n/2}.$$

Back to our toy example

- Consider the case where the target is defined on $\mathbb{R}^n$ and

$$\pi\left(x_{1:n}\right)=\prod_{n=1}^{n}\mathcal{N}\left(x_k;0,1\right),$$
$$\gamma\left(x_{1:n}\right)=\prod_{k=1}^{n}\exp\left(-\frac{x_k^2}{2}\right),\ Z=\left(2\pi\right)^{n/2}.$$

- We select an importance distribution

$$q\left(x_{1:n}\right)=\prod_{k=1}^{n}\mathcal{N}\left(x_k;0,\sigma^2\right).$$

- For SMC, the asymptotic variance is finite only when $\sigma^2 > \frac{1}{2}$ and

$$\frac{\mathbb{V}_{\mathrm{SMC}}\left[\widehat{Z}_n\right]}{Z_n^2} \approx \frac{1}{N}\left[\int \frac{\pi_n^2\left(x_1\right)}{q_1\left(x_1\right)} dx_1 - 1 + \sum_{k=2}^{n} \int \frac{\pi_n^2\left(x_k\right)}{q_k\left(x_k\right)} dx_k - 1\right]$$
$$= \frac{n}{N}\left[\left(\frac{\sigma^4}{2\sigma^2 - 1}\right)^{1/2} - 1\right]$$

compared to

$$\frac{\mathbb{V}_{\mathrm{IS}}\left[\widehat{Z}_n\right]}{Z_n^2} = \frac{1}{N}\left[\left(\frac{\sigma^4}{2\sigma^2 - 1}\right)^{n/2} - 1\right]$$

for SIS.

- For SMC, the asymptotic variance is finite only when $\sigma^2 > \frac{1}{2}$ and

$$\frac{\mathbb{V}_{\text{SMC}}\left[\widehat{Z}_n\right]}{Z_n^2} \approx \frac{1}{N}\left[\int \frac{\pi_n^2(x_1)}{q_1(x_1)}dx_1 - 1 + \sum_{k=2}^{n}\int \frac{\pi_n^2(x_k)}{q_k(x_k)}dx_k - 1\right]$$
$$= \frac{n}{N}\left[\left(\frac{\sigma^4}{2\sigma^2-1}\right)^{1/2} - 1\right]$$

compared to

$$\frac{\mathbb{V}_{\text{IS}}\left[\widehat{Z}_n\right]}{Z_n^2} = \frac{1}{N}\left[\left(\frac{\sigma^4}{2\sigma^2-1}\right)^{n/2} - 1\right]$$

for SIS.

- If select $\sigma^2 = 1.2$ then we saw that it is necessary to employ $N \approx 2 \times 10^{23}$ particles in order to obtain $\frac{\mathbb{V}_{\text{IS}}\left[\widehat{Z}_n\right]}{Z_n^2} = 10^{-2}$ for $n = 1000$.

- For SMC, the asymptotic variance is finite only when $\sigma^2 > \frac{1}{2}$ and

$$\frac{\mathbb{V}_{\text{SMC}}\left[\widehat{Z}_n\right]}{Z_n^2} \approx \frac{1}{N}\left[\int \frac{\pi_n^2\left(x_1\right)}{q_1\left(x_1\right)} dx_1 - 1 + \sum_{k=2}^{n} \int \frac{\pi_n^2\left(x_k\right)}{q_k\left(x_k\right)} dx_k - 1\right]$$
$$= \frac{n}{N}\left[\left(\frac{\sigma^4}{2\sigma^2 - 1}\right)^{1/2} - 1\right]$$

compared to

$$\frac{\mathbb{V}_{\text{IS}}\left[\widehat{Z}_n\right]}{Z_n^2} = \frac{1}{N}\left[\left(\frac{\sigma^4}{2\sigma^2 - 1}\right)^{n/2} - 1\right]$$

for SIS.

- If select $\sigma^2 = 1.2$ then we saw that it is necessary to employ $N \approx 2 \times 10^{23}$ particles in order to obtain $\frac{\mathbb{V}_{\text{IS}}\left[\widehat{Z}_n\right]}{Z_n^2} = 10^{-2}$ for $n = 1000$.
- To obtain the same performance, $\frac{\mathbb{V}_{\text{SMC}}\left[\widehat{Z}_n\right]}{Z_n^2} = 10^{-2}$, SMC requires the use of just $N \approx 10^4$ particles: an improvement by 19 orders of magnitude.

- If you have nice mixing properties, then you can obtain

$$\sigma^2_{SMC}(\varphi) \leq \frac{C}{N}$$

for φ depending only on $X_{n-L+1:n}$.

- If you have nice mixing properties, then you can obtain

$$\sigma^2_{SMC}(\varphi) \leq \frac{C}{N}$$

for φ depending only on $X_{n-L+1:n}$.

- Under the same assumptions, you can also obtain

$$\sigma^2_{SMC} \leq \frac{D.T}{N}.$$

Summary

- Resampling can drastically improve the performance of SIS in models having ‘good’ mixing properties; e.g. state-space models: this can be verified experimentally *and* theoretically.

Summary

- Resampling can drastically improve the performance of SIS in models having 'good' mixing properties; e.g. state-space models: this can be verified experimentally *and* theoretically.
- Resampling does not solve all our problems; only the SMC approximations of the most recent marginals $\pi_n\left(x_{n-L+1:n}\right)$ are reliable; i.e. we can have uniform (in time) convergence bounds.

Summary

- Resampling can drastically improve the performance of SIS in models having 'good' mixing properties; e.g. state-space models: this can be verified experimentally *and* theoretically.
- Resampling does not solve all our problems; only the SMC approximations of the most recent marginals $\pi_n\left(x_{n-L+1:n}\right)$ are reliable; i.e. we can have uniform (in time) convergence bounds.
- The SMC approximation of $\pi_n\left(x_{1:n}\right)$ is only reliable for 'small' n.