# Sequential Monte Carlo methods

## Lecture 16 – SMC samplers

Andreas Svensson, Uppsala University

2017-08-29

UPPSALA
UNIVERSITET

# Outline – Lecture 3

**Aim:** Introduce the SMC sampler

**Outline:**

1. Problem formulation – when is an SMC sampler applicable?
2. The annealing/tempering idea
3. Constructing the SMC sampler
4. User aspects

# Main reference

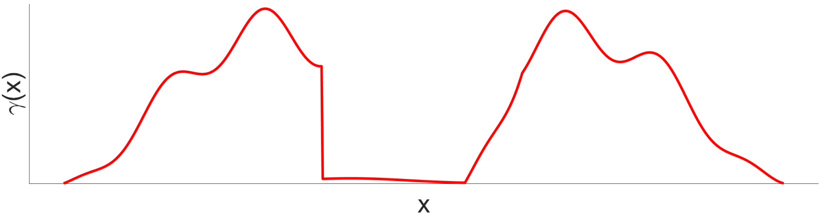
Pierre Del Moral, Arnaud Doucet, and Ajay Jasra 2006:. **Sequential Monte Carlo samplers.** *Journal of the Royal Statistical Society: Series B*, 68(3), pp. 411–436.

# Problem formulation

Let $\mathcal{X}$ be a space on which a probability density $\gamma$ is defined. Let $\widetilde{\gamma}$ be an unnormalized version of the density, as $\gamma(x) = \frac{\widetilde{\gamma}(x)}{Z}$. Assume that only $\widetilde{\gamma}(x)$ can be evaluated pointwise.

**Goal:** Generate $N$ samples $x^i \in \mathcal{X}$ from the density $\gamma(x)$.

(A typical situation is when $\gamma$ is a parameter posterior or $Z$ is the marginal likelihood/model evidence.)



# Most common solution

MCMC?

SMC sampler is an alternative!

# Metropolis-Hastings

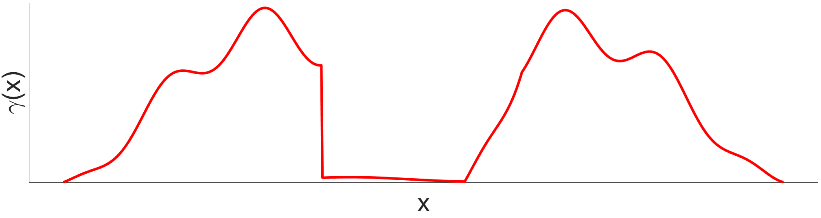
$\gamma$-**invariant Metropolis-Hastings** (a reminder)

**for k = 1, . . .**

Propose a new sample $x'$ from a proposal $r(x' \mid x_k)$

Compute acceptance rate $\alpha = \max(1, \frac{\gamma(x')}{\gamma(x_k)} \frac{r(x' \mid x_k)}{r(x_k \mid x')})$

Set $x_{k+1} \leftarrow x'$ with probability $\alpha$, otherwise set $x_{k+1} \leftarrow x_k$

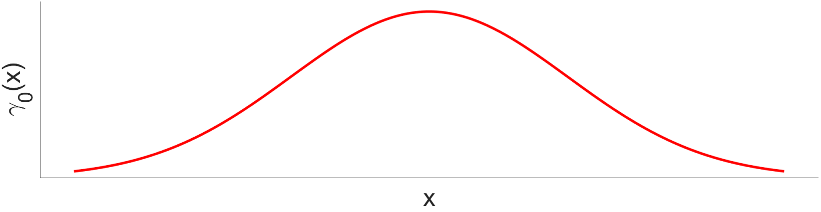


# Annealing/tempering

Sequential Monte Carlo needs something **sequential**. Construct a sequence which transitions 'smoothly' in $K$ steps from a simple initial $\gamma_0(x)$ to the sought $\gamma_K(x) \equiv \gamma(x)$.

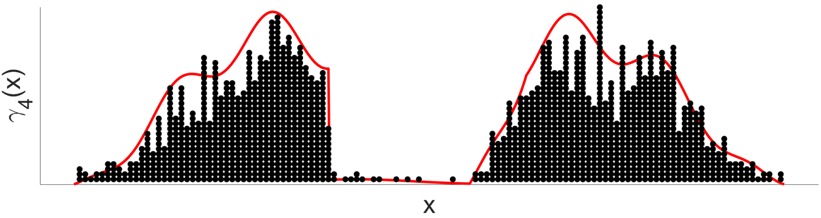
**For example:**

- If $\gamma(x)$ is a posterior $\gamma(x) \propto p(y \mid x)p(x)$, then $\gamma_k(x) \propto p(y \mid x)^{\tau_k} p(x)$, $\tau_k = k/K$ (likelihood tempering)
- If $\gamma(x)$ depends on some data $y_{1:K}$ as $\gamma(x) = p(x \mid y_{1:K})$, then $\gamma_k(x) = p(x \mid y_{1:k})$ (data tempering)



# The SMC sampler: a sneak peek

**Intuition:** Track the evolving sequence $\gamma_0, \gamma_1, \ldots, \gamma_K$ using a weighting - resampling - propagation scheme.



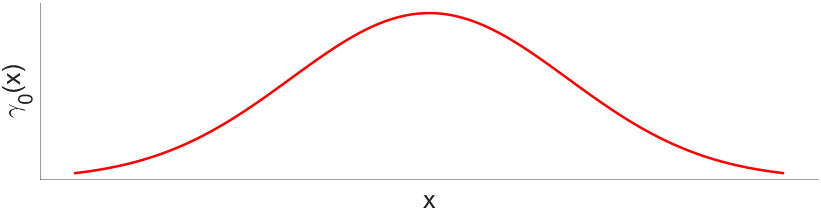

**How do we do it?** This sequence (unlike the state inference problem) is not defined as a state-space model, neither does it fall into the general SMC formulation (yet).

# Attempt I

*Let's try to make use of the sequence* $\gamma_0, \gamma_1, \ldots, \gamma_K$*:*

Sample $x^i$ from $\gamma_0$ (we assume we can do that)
**for** $k = 1$ **to** $K$
    Evaluate $\widetilde{w}_k^i(x^i) = \frac{\gamma_k(x^i)}{\gamma_{k-1}(x^i)}$
Set weights $\widetilde{w}^i \propto \prod_{k=1}^K \widetilde{w}_k^i$ and normalize

Valid but inefficient: effectively importance sampling with proposal $\gamma_0$ and target $\gamma$.



# Attempt II

Sample $x_0^i$ from $\gamma_0$ (we assume we can do that)

**for** $k = 1$ **to** $K$

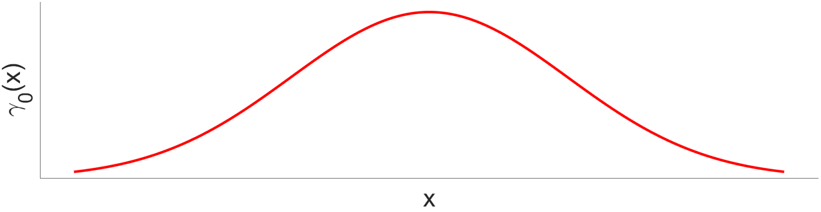
Use some Markov kernel $\kappa_k$ to sample new $x_k^i$ from $\kappa_k(x_{k-1}^i, x_k^i)$

Set weights $w_k^i \propto w_{k-1}^i \frac{\gamma_k(x_k^i)}{\eta_k(x_k^i)}$ and normalize

If needed: Resample and set weights $w_k^i = 1/N$

$\eta_k(x_k^i)$ is $\eta_k(x_k) = \int_{\mathcal{X}^k} \gamma_0(x_0) \prod_{k=1}^{k} K_k(x_{k-1}, x_k) dx_{0:k-1}$.

In most cases intractable



# Attempt III

Introduce a backward kernel $\lambda_{k-1}(x_k, x_{k-1})$, and define

$$\pi_0(x_0) = \gamma_0(x_0), x_0 \in \mathcal{X}$$

$$\pi_1(x_{0:1}) = \gamma_1(x_1)\lambda_0(x_1, x_0), x_{0:1} \in \mathcal{X} \times \mathcal{X} = \mathcal{X}^2$$

$$\vdots$$

$$\pi_K(x_{0:K}) = \gamma_p(x_p) \prod_{k=1}^{K} \lambda_{k-1}(x_k, x_{k-1}), x_{0:K} \in \mathcal{X}^{K+1}$$

- $\gamma_k$ is defined on $\mathcal{X}$, whereas $\pi_k$ is defined on $\mathcal{X}^{k+1}$
- The marginal with index $k$ of $\pi_k(x_{0:k})$ is $\int \pi_k(x_{0:k}) dx_{0:k-1} = \gamma_k(x_k)$
- The marginal with index $j < k$ of $\pi_k(x_{0:k})$ is $\int \pi_k(x_{0:k}) dx_{0:j-1,j+1:k} \neq \gamma_j(x_j)$

# Attempt III

We now have a sequence $\pi_0, \pi_1, \dots \pi_p$ (with marginals that we are interested in) defined on the spaces $\mathcal{X}, \mathcal{X}^2, \dots, \mathcal{X}^{K+1}$.

Use the general SMC scheme!

Sample $x^i$ from $\gamma_0$ and set weights $w_0^i = 1/N$
**for** $k = 1$ **to** $K$
  Use $\gamma_k$-invariant Markov kernel $\kappa_k$ to sample $x_k^i$
  Set weights $\widetilde{w}_k^i = w_{k-1}^i \omega(x_{k-1:k}^i)$ and normalize to $w_k^i$
  If ESS too low, resample and set $w_k^i = 1/N$

Here, $\omega(x_{k-1:k}) = \frac{\gamma_k(x_k)\lambda_{k-1}(x_k, x_{k-1})}{\gamma_{k-1}(x_{k-1})\kappa_k(x_{k-1}, x_k)}$

# Attempt III

How to choose the artificial backward kernels $\lambda_{k-1}$?

- To minimize the variance of the weights, the optimal choice is $\lambda_{k-1} = \frac{\eta_{k-1}\kappa_k}{\eta_k}$
  $\Rightarrow$ The intractable IS solution
- Keeping variance low *and* the weights tractable:
  A generic choice, if $\kappa_k$ has invariant distribution $\gamma_k$, is

$$\lambda_{k-1}(x_k, x_{k-1}) = \frac{\gamma_k(x_{k-1})\kappa_k(x_{k-1}, x_k)}{\gamma_k(x_k)}$$

$\Rightarrow$

$$\omega(x_{k-1:k}) = \frac{\gamma_k(x_{k-1})}{\gamma_{k-1}(x_{k-1})}$$

# Attempt III

Design choices made: $\kappa_k$ and $\lambda_{k-1}$

The $\gamma_k$-invariant Markov kernel $\kappa_k$ is **one option** for propagating the samples, $q_k$ in the general SMC framework.

The backward kernel $\lambda_{k-1}$ is **part of the model specification** of $\pi_0, \ldots, \pi_K$, in the SMC context. (But since we are only interested in a marginal of $\pi_K$ not depending on $\lambda_{k-1}$, it may appear to be part of the inference rather than the model.)

# The final generic SMC sampler

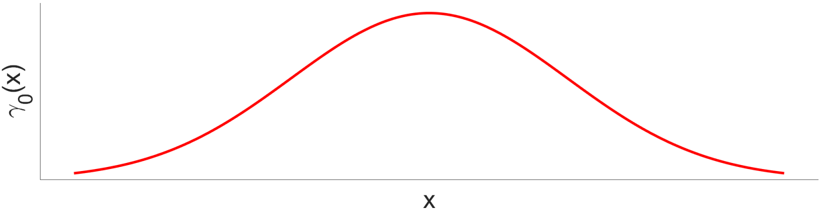
Sample $x^i$ from $\gamma_0$ (we assume we can do that) and set weights $w_0^i = 1/N$

**for** $k = 1$ **to** $K$

Set weights $w_k^i \propto w_{k-1}^i \frac{\gamma_k(x_{k-1}^i)}{\gamma_{k-1}(x_{k-1}^i)}$ and normalize

If ESS too low, resample and set $w_k^i = 1/N$

Use $\gamma_k$-invariant Metropolis-Hastings to sample $x_k^i$



# Estimating $Z$

For notational convenience, we have implicitly assumed we can evaluate $\gamma(x)$ exactly for any $x \in \mathcal{X}$. The SMC is also applicable if we only can evaluate $\widetilde{\gamma}(x)$, where $\gamma(x) = \frac{\widetilde{\gamma}(x)}{Z}$.

If $Z = Z_K$ is of interest, we can estimate $Z_K/Z_0$ as

$$\widehat{\frac{Z_K}{Z_0}} = \prod_{k=1}^{K} \widehat{\frac{Z_k - 1}{Z_k}}$$

where

$$\widehat{\frac{Z_{k-1}}{Z_k}} = \sum_{i=1}^{N} w_{k-1}^i \omega(x_{k-1:k}^i)$$

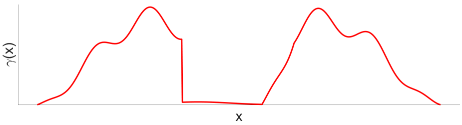
and $Z_0$ is the normalizing constant of the user-chosen $\gamma_0$.

(Superior to annealed importance sampling.)

# SMC sampler vs MCMC

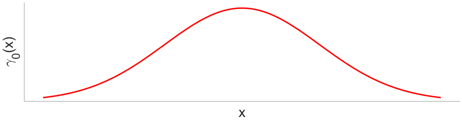
## MCMC (Metropolis-Hastings)

Set initial $x_0$
**for k = 1, . . .**
Propose a new sample $x'$ from $r(x' \mid x_k)$
Compute $\alpha = \max(1, \frac{\gamma(x')}{pi(x_k)} \frac{q(x' \mid x_k)}{q(x_k \mid x')})$
Set $x_{k+1} \leftarrow x'$ with probability $\alpha$,
otherwise $x_{k+1} \leftarrow x_k$
**end**




## SMC sampler

Sample $x^i$ from $\gamma_0$ and set weights $w_0^i = 1/N$
**for** $k = 1$ **to** $K$
Set $\widetilde{w}_k^i = w_{k-1}^i \frac{\gamma_k(x_{k-1}^i)}{\gamma_{k-1}(x_{k-1}^i)}$ and normalize
If ESS too low, resample and set $w_k^i = 1/N$
Sample $x_k^i$ by $\gamma_k$-invariant Metropolis-Hastings
**end**



# User's generic recipe

1. Design a simulated annealing sequence (e.g., likelihood or data tempering)
2. Design a MCMC kernel (typically Metropolis-Hastings) for $\gamma_k$
3. Run the SMC sampler

# Automatic adaptation

- Adaptation of the MCMC kernels

  Paul Fearnhead and Benjamin M. Taylor 2013:. **An adaptive sequential Monte Carlo sampler.** *Bayesian analysis*, 8(2), pp. 411–438.

- Adaptation of the tempering sequence

  Yan Zhou, Adam M. Johansen and John A.D. Aston 2016. **Toward Automatic Model Comparison: An Adaptive Sequential Monte Carlo Approach.** *Journal of Computational and Graphical Statistics*, 25(3), pp. 701–726.

# Some further developments

- Approximate Bayesian computations (ABC)

  Pierre Del Moral, Arnaud Doucet, Ajay Jasra 2012:. **An adaptive sequential Monte Carlo method for approximate Bayesian computation.** *Statistics and computing*, 22(5), pp. 1009–1020.

- Use SMC sampler for unknown parameters in a state-space model

  Nicolas Chopin, Pierre E. Jacob, Omiros Papasiliopoulos 2013:. **SMC$^2$: An efficient algorithm for sequential analysis of state space models.** *Journal of the Royal Statistical Society: Series B*, 75(3), pp. 397–426.

# A few concepts to summarize lecture 16

- SMC sampler is an alternative to MCMC
- The simulated annealing sequence is key
- The formal construction is made possible by the use of backward kernels $L_k$
- Loosely speaking, the SMC sampler manages (propagates, terminates and duplicates) $N$ parallel MCMC chains