Sequential Monte Carlo methods

Lecture 16 – SMC samplers

Fredrik Lindsten, Linköping University

2019-08-30

Summary of day 4

Simulate a Markov chain which is designed in such a way that its stationary distribution coincides with the target distribution.

An MCMC sampler generates the Markov chain $\{x[m]\}_{m=1}^{M}$ by:

- **Initialize:** set $x[1]$ arbitrarily.
- **For** $m = 2$ **to** M**:** sample $x[m] \sim \kappa(x[m-1], x^\star)$.

$\kappa(x, x^\star)$ is a **Markov kernel** on $\mathcal{X}$, i.e. a conditional distribution for the next state $x^\star$ given the current state x.

Summary of day 4

Algorithm 1 Pseudo-marginal Metropolis Hastings

1. **Initialize ($m = 1$):** Set $\theta[1]$ and run a particle filter for $\hat{z}[1]$.
2. **For $m = 2$ to M, iterate:**
   a. Sample $\theta' \sim q(\theta \mid \theta[m-1])$.
   b. Sample $\hat{z}' \sim \psi(z \mid \theta', y_{1:T})$ (i.e. run a particle filter).
   c. With probability
   $$\alpha = \min\left(1, \frac{\hat{z}' p(\theta')}{\hat{z}[m-1] p(\theta[m-1])} \frac{q(\theta[m-1] \mid \theta')}{q(\theta' \mid \theta[m-1])}\right)$$
   set $\{\theta[m], \hat{z}[m]\} \leftarrow \{\theta', \hat{z}'\}$ (accept candidate sample) and with prob. $1 - \alpha$ set $\{\theta[m], \hat{z}[m]\} \leftarrow \{\theta[m-1], \hat{z}[m-1]\}$ (reject candidate sample).

Summary of day 4

Particle Gibbs kernel: A Markov kernel $\kappa_{N,\theta}(x_{0:T}, x^\star_{0:T})$ on $\mathcal{X}^{T+1}$.

Particle Gibbs: Run a particle filter, but at each time step

- sample only $N-1$ particles in the standard way.
- set the Nth particle deterministically: $x_t^N = x_t$ and $a_t^N = N$.
- At final time $t = T$, output $x^\star_{0:T} = x^b_{0:T}$ with $b \sim \mathcal{C}(\{w_T^i\}_{i=1}^N)$

Ancestor sampling: Replace "set $a_t^N = N$" with: sample $a_t^N \in \{1, \ldots, N\}$ with

$$\mathbb{P}(A_t^N = j) \propto w_{t-1}^j p(x_t \mid x_{t-1}^j, \theta).$$

Can **significantly** improve mixing for small N/large T.

Summary of day 4

SMC can be used to approximate a **sequence** of probability distributions $\{\pi_k(x_{0:k})\}_{k\geq 0}$ on a sequence of probability spaces of **increasing dimension**, $\mathcal{X}_{0:k} = \mathcal{X}_{0:k-1} \times \mathcal{X}_k$.

At iteration k**:** given $\{x^i_{0:k-1}, w^i_{k-1}\}_{i=1}^N$

Resampling: Sample a^i_k with $\mathbb{P}(a^i_k = j) = \nu^j_{k-1}$, $j = 1, \ldots, N$.

Propagation: $x^i_k \sim q_k(x_k \mid x^{a^i_k}_{1:k-1})$ and $x^i_{0:k} = (x^{a^i_k}_{0:k-1}, x^i_k)$.

Weighting: $$w^i_k \propto \frac{w^{a^i_k}_{k-1}}{\nu^{a^i_k}_{k-1}} \frac{\widetilde{\pi}_k(x^i_{0:k})}{\widetilde{\pi}_{k-1}(x^{a^i_k}_{0:k-1}) q_k(x^i_k \mid x^{a^i_k}_{0:k-1})}.$$

Outline – Lecture 16

Aim: See how we can use SMC for inference even in the absence of any sequential structure in the model.

Outline:

1. Problem formulation
2. The annealing/tempering idea
3. Constructing the "SMC sampler"
4. User aspects

Main reference

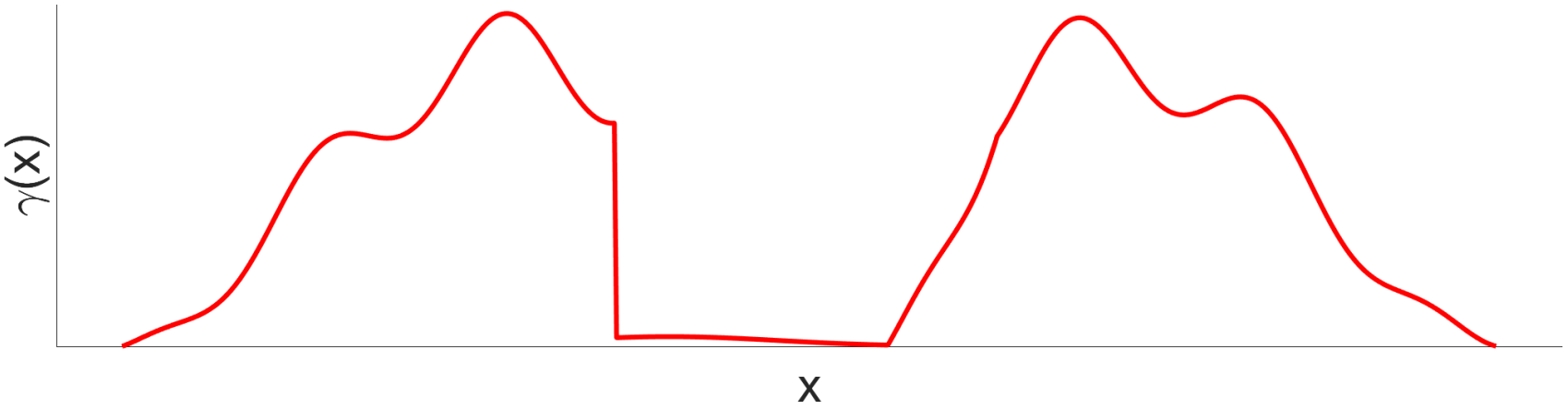
Pierre Del Moral, Arnaud Doucet, and Ajay Jasra 2006:. **Sequential Monte Carlo samplers.** *Journal of the Royal Statistical Society: Series B*, 68(3), pp. 411–436.

Problem formulation

Let $\mathcal{X}$ be a space on which a probability density γ 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 $\gamma(x)$.

ex) Typical situation: $\gamma(x) = p(x|y)$, $\widetilde{\gamma}(x) = p(x, y)$ and $Z = p(y)$.



Most common solution

MCMC?

SMC sampler is an alternative!

Metropolis-Hastings

Metropolis–Hastings targeting γ (a reminder)

for k = 1, 2, . . .

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

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

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

Annealing/tempering

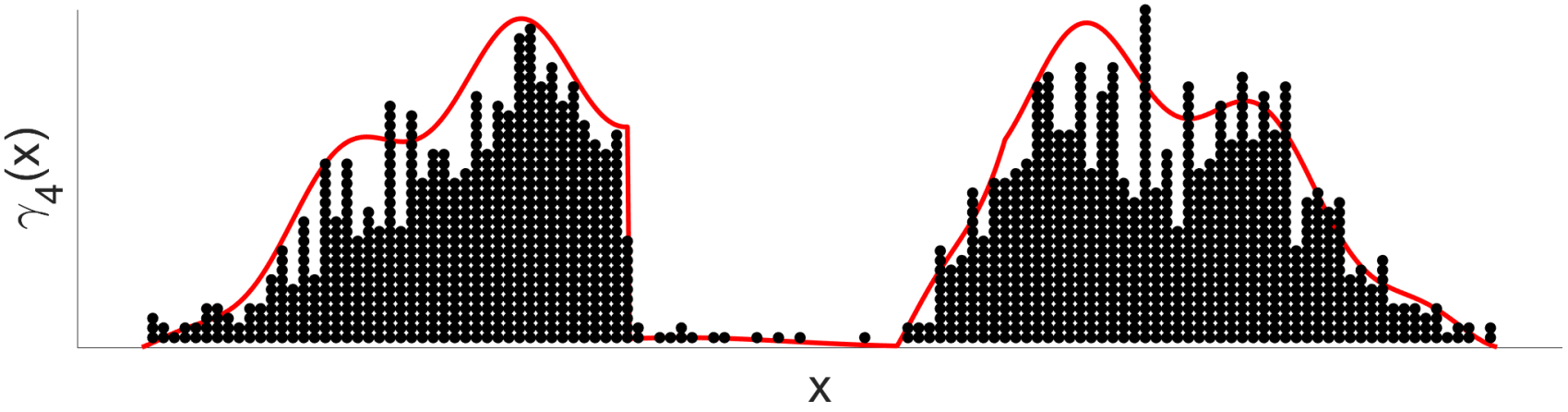
Sequential Monte Carlo needs something **sequential**. Construct a sequence which transitions 'smoothly' in K steps from a simple initial $\gamma_0(x)$ to the sought $\gamma_K(x) \equiv \gamma(x)$.

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).

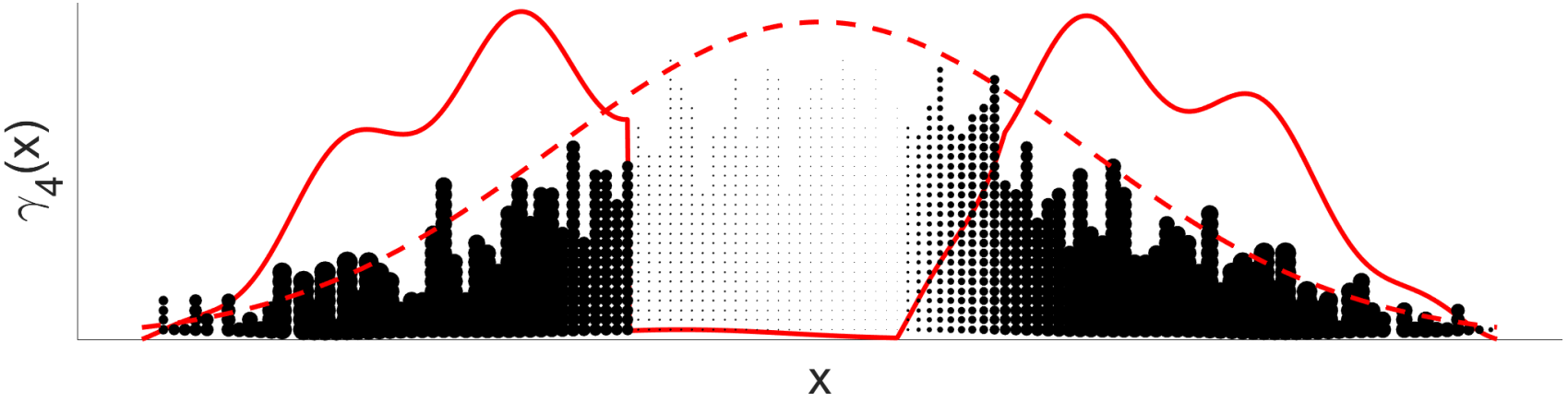
Attempt I

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

Sample x^i from γ_0 and set $\widetilde{w}_0^i(x^i) = 1$

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

Evaluate $\widetilde{w}_k^i(x^i) = \frac{\gamma_k(x^i)}{\gamma_{k-1}(x^i)} \widetilde{w}_{k-1}^i$



Valid but inefficient: effectively importance sampling with proposal γ_0 and target γ.

Attempt II

Sample x_0^i from γ_0

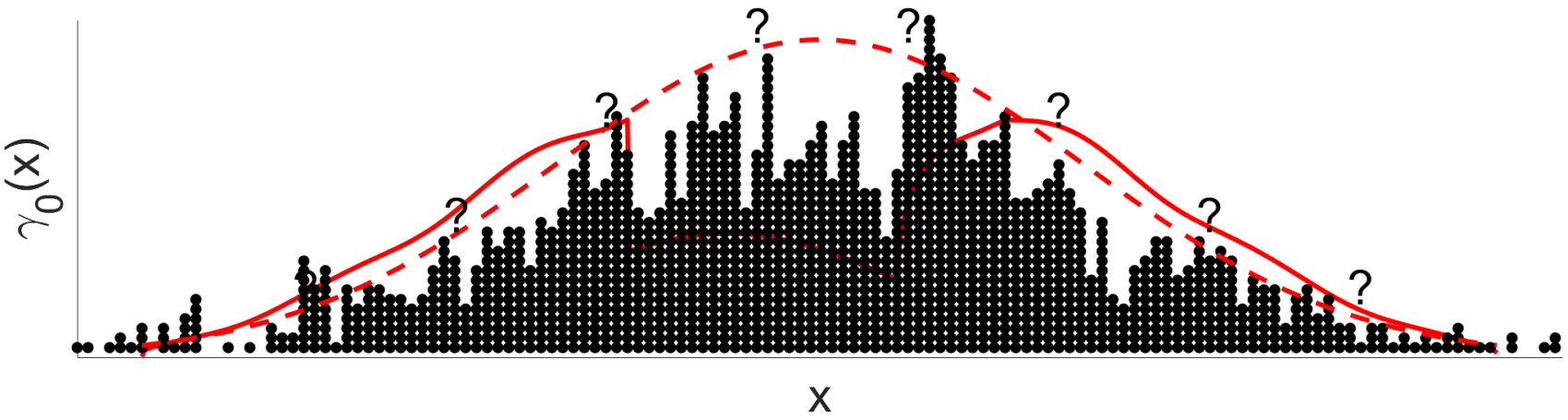
for $k = 1$ **to** K

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

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

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

In most cases intractable



General SMC

Recall:

SMC can be used to approximate a **sequence** of probability distributions

$$\{\pi_k(x_{0:k})\}_{k\geq 0}$$

on a sequence of probability spaces of **increasing dimension**,

$$\mathcal{X}_{0:k} = \mathcal{X}_{0:k-1} \times \mathcal{X}_k.$$

- The intermediate target distributions can be chosen arbitrarily.
- We need to be able to recover the original distribution of interest (here $\gamma_K(x) \equiv \gamma(x)$) at iteration K.

Attempt III

We have a **sequence of distributions** $\gamma_0(x)$, $\gamma_1(x)$, $\ldots$, $\gamma_K(x)$ evolving smoothly from something simple (γ_0) to the target distribution of interest (γ_K).

Problem with directly applying SMC to this sequence: The γ_k's are all defined on the same space $\mathcal{X}$.

SMC requires spaces of **increasing dimension**.

Attempt III

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

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

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

$$\vdots$$

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

- γ_k is defined on $\mathcal{X}$, whereas π_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 $\pi_0, \pi_1, \ldots \pi_K$ (with marginals that we are interested in) defined on the spaces $\mathcal{X}, \mathcal{X}^2, \ldots, \mathcal{X}^{K+1}$.

Use the general SMC scheme!

Sample $x^i \sim \gamma_0(x_0)$ and set weights $w_0^i = 1/N$
for $k = 1$ **to** K
    If ESS too low, resample and set $w_k^i = 1/N$
    Use Markov kernel κ_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

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 do we choose the backward kernels λ_{k-1}?

Assume that κ_k is an **MCMC kernel** with stationary distribution γ_k. We can then select λ_{k-1} as its **reversal**:

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

$\Rightarrow$

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

Attempt III

Design choices made: κ_k and λ_{k-1}

The γ_k-invariant Markov kernel κ_k is **one option** for propagating the samples, q_k in the general SMC framework.

The backward kernel λ_{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 π_K not depending on λ_{k-1}, it may appear to be part of the inference algorithm rather than the model.)

The final SMC sampler

Sample x_0^i from γ_0 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 $w_k^i = 1/N$

Sample x_k^i from MCMC kernel with stationary distribution γ_k.

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

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

and Z_0 is the normalizing constant of the user-chosen γ_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 $\alpha = \min(1, \frac{\gamma(x')}{\gamma(x_k)} \frac{r(x_k \mid x')}{r(x' \mid x_k)})$
    Set $x_{k+1} \leftarrow x'$ with probability α,
        otherwise $x_{k+1} \leftarrow x_k$
end

SMC sampler

Sample x_0^i from γ_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 γ_k-invariant Metropolis-Hastings
end

User's generic recipe

1. Design a simulated annealing sequence (e.g., likelihood or data tempering)
2. Design MCMC kernel κ_k (typically Metropolis-Hastings) for γ_k
3. Design backward kernel λ_{k-1}. Simplest choice is as the reversal of κ_k, but other options are available.
4. 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 λ_k