Stat 535 C - Statistical Computing & Monte Carlo Methods

Lecture 23 - 4th April 2006

Arnaud Doucet

Email: arnaud@cs.ubc.ca

1.1– Outline

- Review of Sequential Monte Carlo.
- Sequential Monte Carlo for Static Problems.
- Limitations.

2.1– Review of SMC Methods

- The SMC approach can be used to sample from any sequence of target distributions

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

- Standard application:

$$\pi_n\left(x_{1:n}\right) \propto \mu\left(x_1\right) \prod_{k=2}^{n} f\left(x_k \middle| x_{k-1}\right) \prod_{k=1}^{n} g\left(y_k \middle| x_k\right).$$

- Rao-blackwellisation examples:

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

- We use the IS identities

$$\pi_n(x_{1:n}) = \frac{w_n(x_{1:n})\, q_n(x_{1:n})}{Z_n},$$

$$Z_n = \int w_n(x_{1:n})\, q_n(x_{1:n})\, dx_{1:n}$$

where

$$w_n(x_{1:n}) = \frac{\gamma_n(x_{1:n})}{q_n(x_{1:n})} \propto \frac{\pi_n(x_{1:n})}{q_n(x_{1:n})}$$

- If $X_{1:n}^{(i)} \sim q_n(x_{1:n})$ then

$$q_n^N(x_{1:n}) = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_{1:n}^{(i)}}(x_{1:n})$$

- By plugging $q_n^N(x_{1:n})$ in place of $q_n(x_{1:n})$

$$\pi_n^N(x_{1:n}) = \sum_{i=1}^{N} W_n^{(i)} \delta_{X_{1:n}^{(i)}}(x_{1:n})$$

where

$$W_n^{(i)} \propto w_n\left(X_{1:n}^{(i)}\right), \ \sum_{i=1}^{N} W_n^{(i)} = 1$$

and

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

- We use

$$q_n\left(x_{1:n}\right) = q_1\left(x_1\right)\prod_{k=2}^{n} q_k\left(x_k\middle|\, x_{1:k-1}\right),$$

$$w_n\left(x_{1:n}\right) = w_n\left(x_{1:n-1}\right)\frac{\gamma_n\left(x_{1:n}\right)}{\gamma_{n-1}\left(x_{1:n-1}\right) q_n\left(x_n\middle|\, x_{1:n-1}\right)}.$$

- The variance of $\left\{w_n\left(X_{1:n}^{(i)}\right)\right\}$ tends to increase so when the ESS is too low we resample the particle approximation $\left\{W_n^{(i)}, X_{1:n}^{(i)}\right\}$ to obtain $\left\{1/N, X_{1:n}^{(i)}\right\}$.

2.1– Review of SMC Methods

At time $n-1$, $\left\{W_{n-1}^{(i)}, X_{1:n-1}^{(i)}\right\}$

- *Sampling Step.* For $i = 1, ..., N$, sample $X_n^{(i)} \sim q_n\left(\cdot \middle| X_{1:n-1}^{(i)}\right)$

$$W_n^{(i)} \propto W_{n-1}^{(i)} \frac{\gamma_n\left(X_{1:n}^{(i)}\right)}{\gamma_{n-1}\left(X_{1:n-1}^{(i)}\right) q_n\left(x_n \middle| X_{1:n-1}^{(i)}\right)}$$

- *Resampling Step.* If variance of weights $\left\{W_n^{(i)}\right\}$ high, resample $\left\{W_n^{(i)}, X_{1:n}^{(i)}\right\}$

to obtain $\left\{N^{-1}, X_{1:n}^{(i)}\right\}$.

- An approximation of π_n is given by

$$\pi_n^N\left(x_{1:n}\right) = \sum_{i=1}^{N} W_n^{(i)} \delta_{X_{1:n}^{(i)}}\left(x_{1:n}\right)$$

- An approximation of

$$\frac{Z_n}{Z_{n-1}} = \frac{\int \gamma_n\left(x_{1:n}\right) dx_{1:n}}{\int \gamma_{n-1}\left(x_{1:n-1}\right) dx_{1:n-1}}$$

is given by

$$\widehat{\frac{Z_n}{Z_{n-1}}} = \sum_{i=1}^{N} W_{n-1}^{(i)} \frac{\gamma_n\left(X_{1:n}^{(i)}\right)}{\gamma_{n-1}\left(X_{1:n-1}^{(i)}\right) q_n\left(x_n \middle| X_{1:n-1}^{(i)}\right)}$$

2.2– Limitations

• One cannot hope to estimate with a fixed precision a target distribution of increasing dimension.

• Aat best, we can expect results of the following form

$$E\left[\left(\int \varphi\left(x_{n-L+1:n}\right)\left(\pi_n^N\left(dx_{n-L+1:n}\right)-\pi_n\left(dx_{n-L+1:n}\right)\right)\right)^2\right] \leq \frac{C_L\left\|\varphi\right\|}{N}$$

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

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

with $0 \leq \lambda < 1$.

• Under such assumptions, there is no accumulation of errors over time.

2.2– Limitations

• SMC are very useful for dynamic models and more generally to estimate any sequence of distributions whose dimension increases over time.

• Unfortunately, many important problems do NOT fit into this framework. How do you estimate a fixed target distribution $\pi(x)$ using SMC??

• We need to develop more elaborate methods for such problems.

3.1– Objectives

• Let $\{\pi_n\}_{n\geq 1}$ be a *sequence of probability distributions* defined on E such that $\pi_n(dx) = \pi_n(x)\,dx$ and each $\pi_n(x)$ is known *up to a normalizing constant*, i.e.

$$\pi_n(x) = \underbrace{Z_n^{-1}}_{\text{unknown}} . \underbrace{\gamma_n(x)}_{\text{known}}.$$

• Estimate expectations $\int \varphi(x)\,\pi_n(dx)$ and/or normalizing constants Z_n *sequentially*; i.e. first π_1 then π_2 and so on.

• *Objectives*: Obtain SMC (sampling/resampling population-based) algorithms to solve this problem.

• Standard SMC methods do not apply.

- *Sequential Bayesian Inference:* $\pi_n(x) = p(x|y_{1:n})$.
- *Global optimization:* $\pi_n(x) \propto [\pi(x)]^{\eta_n}$ with $\{\eta_n\}$ increasing sequence such that $\eta_n \to \infty$.
- *Sampling from a fixed target* π : $\pi_n(x) \propto [\mu_1(x)]^{\eta_n} [\pi(x)]^{1-\eta_n}$ where μ_1 easy to sample and $\eta_1 = 1$, $\eta_n < \eta_{n-1}$ and $\eta_P = 0$.
- *Rare event simulation* $\pi(A) \ll 1$: $\pi_n(x) \propto \pi(x) 1_{E_n}(x)$ with Z_1 known, $E_1 = E$, $E_n \subset E_{n-1}$ and $E_P = A$ then $Z_P = \pi(A)$.

3.3– Brief Review of Standard Importance Sampling

• Let the *target distribution* be $\pi_k(x) = Z_k^{-1}\gamma_k(x)$ and μ_k be a so-called *importance distribution* then

$$\pi_k(x) = \frac{w_k(x)\,\mu_k(x)}{\int w_k(x)\,\mu_k(x)\,dx} \text{ where } w_k(x) = \frac{\gamma_k(x)}{\mu_k(x)},$$

$$Z_k = \int w_k(x)\,\mu_k(x)\,dx$$

• By sampling N i.i.d. particles $X_k^{(i)} \sim \mu_k$ then $\widehat{\mu}_k(dx) = \frac{1}{N}\sum_{i=1}^N \delta_{X_k^{(i)}}(dx)$ and

$$\widehat{\pi}_k(dx) = \sum_{i=1}^N W_k^{(i)}\delta_{X_k^{(i)}}(dx) \text{ where } W_k^{(i)} \propto w_k\left(X_k^{(i)}\right),\ \sum_{i=1}^N W_k^{(i)} = 1,$$

$$\widehat{Z}_k = \frac{1}{N}\sum_{i=1}^N w_k\left(X_k^{(i)}\right).$$

3.4– Problems associated to Importance Sampling

• Importance Sampling (IS) is a straightforward method to use if μ_k is easy to sample.

• Under weak assumptions, we can obtain asymptotically consistent estimates of $\int \varphi(x) \widehat{\pi}_k(dx)$ and $\widehat{Z}_k$.

.... so why do people use MCMC in 99.99% of cases???

• For the estimates to have reasonable variances (if they exist!), one needs to select very carefully the importance distribution.

• To compute $\int \varphi(x) \pi_k(dx)$ by IS, the optimal distribution depends on φ but in statistics we often simply want μ_k as "close" to π_k as possible.

• For problems routinely addressed in statistics, this is very difficult.

3.5– What we propose to do

• "Philosophy": Start by doing simple things before trying to do complex things; same idea used in simulated annealing, simulated tempering etc.

• Develop a sequential/iterative IS strategy where we start by approximating a simple target distribution π_1. Then targets evolve over time and *we build the importance distribution sequentially.*

• At time n, we use μ_{n-1} to build μ_n.

• This approach makes sense if the sequence $\{\pi_n\}$ is not arbitrary; i.e. π_{n-1} somewhat close to π_n.

3.6– Initialization via Standard Importance Sampling

• At time 1, sample N ($N \gg 1$) particles $X_1^{(i)} \sim \mu_1$ to obtain the following IS estimates

$$\widehat{\pi}_1(dx) = \sum_{i=1}^{N} W_1^{(i)} \delta_{X_1^{(i)}}(dx) \text{ where } W_1^{(i)} \propto w_1\left(X_1^{(i)}\right), \sum_{i=1}^{N} W_1^{(i)} = 1,$$

$$\widehat{Z}_1 = \frac{1}{N} \sum_{i=1}^{N} w_1\left(X_1^{(i)}\right)$$

• *Remark*: Estimates have reasonable variance only if discrepancy between π_1 and μ_1 small; hence the need to start with easy to sample or approximate π_1.

3.7– Moving Forward

- At time $n-1$, one has N particles $\left\{X_{n-1}^{(i)}, W_{n-1}^{(i)}\right\}$

$$X_{n-1}^{(i)} \sim \mu_{n-1} \text{ and } W_{n-1}^{(i)} \propto \frac{\pi_{n-1}\left(X_{n-1}^{(i)}\right)}{\mu_{n-1}\left(X_{n-1}^{(i)}\right)}.$$

- Move the particles according to transition kernel

$$X_n^{(i)} \sim K_n\left(X_{n-1}^{(i)}, \cdot\right) \Rightarrow \mu_n\left(x'\right) = \int \mu_{n-1}\left(x\right) K_n\left(x, x'\right) dx$$

- Optimal transition kernel $K_n\left(x, x'\right) = \pi_n\left(x'\right)$ cannot be used so we need alternatives.

3.8– Transition kernels

- $K_n(x, x') = K_n(x')$ with
  - simple parametric form (e.g. Gaussian, multinomial etc.);
  - semi-parametric based on $\widehat{\mu}_{n-1}(dx)$ (e.g. West, 1993; Titterington, 2001)

  complexity $O\left(N^2\right)$.

- $K_n(x, x')$ MCMC kernel of invariant distribution π_n.
  - burn-in correction by importance sampling.
  - scaling of proposal can depend on $\left\{X_{n-1}^{(i)}\right\}$ (Crisan & D., 2000 Chopin, 2002)

- $K_n(x, x')$ approximation of a Gibbs sampler of invariant distribution π_n.

nitialization; $n = 1$.

For $i = 1, ..., N$, sample $X_1^{(i)} \sim \mu_1(\cdot)$ and set

$$w_1\left(X_1^{(i)}\right) = \frac{\gamma_1\left(X_1^{(i)}\right)}{\mu_n\left(X_1^{(i)}\right)},\ W_1^{(i)} \propto w_1\left(X_1^{(i)}\right).$$

At time n; $n \geq 1$.

For $i = 1, ..., N$, sample $X_n^{(i)} \sim K_n\left(X_{n-1}^{(i)}, \cdot\right)$ and set

$$w_n\left(X_n^{(i)}\right) = \frac{\gamma_n\left(X_n^{(i)}\right)}{\mu_n\left(X_n^{(i)}\right)},\ W_n^{(i)} \propto w_n\left(X_n^{(i)}\right)$$

$$\text{where } \mu_n(x_n) = \int \mu_{n-1}(dx_{n-1})\, K_n(x_{n-1}, x_n).$$

3.10– Limitations of this approach

- In most cases, we *cannot* compute the marginal importance distribution

$$\mu_n(x_n) = \int \mu_{n-1}(x_{n-1}) K_n(x_{n-1}, x_n) dx_{n-1}$$

$$= \int \mu_1(x_1) \prod_{k=2}^{n} K_k(x_{k-1}, x_k) dx_{1:n-1}.$$

$\Rightarrow$ Hence we cannot use Importance Sampling.

3.10– Limitations of this approach

- Monte Carlo approximation

$$\widehat{\mu}_n(x_n) = \int \widehat{\mu}_{n-1}(dx_{n-1})\, K_n(x_{n-1}, x_n) = \frac{1}{N}\sum_{i=1}^{N} K_n\left(X_{n-1}^{(i)}, x_n\right).$$

$\rightsquigarrow$ Computationally intensive $O\left(N^2\right)$.

$\rightsquigarrow$ Impossible if $K_n(x, x')$ cannot be evaluated pointwise;

e.g. Metropolis-Hastings kernel where

$$K_n(x, dx') = \alpha(x, x')\, q(x, dx') + \underbrace{\left(1 - \int \alpha(x, u)\, q(x, du)\right)}_{\text{unknown}} \delta_x(dx')$$

3.11– How to use local moves

• *Problem summary*: It is impossible to compute pointwise $\mu_n(x_n)$ hence $\gamma_n(x_n)/\mu_n(x_n)$ except when $n = 1$.

• *Solution*: Perform importance sampling on extended space.

  • At time 2,

$$\frac{\pi_2(x_2)}{\mu_2(x_2)} = \frac{\pi_2(x_2)}{\int \mu_1(dx_1) K_2(x_1, x_2)} \text{ cannot be evaluated}$$

but alternative weights can be defined

$$\frac{\text{new joint target distribution}}{\text{joint importance distribution}} = \frac{\pi_2(x_2) L_1(x_2, x_1)}{\mu_1(x_1) K_2(x_1, x_2)}$$

where $L_1(x_2, x_1)$ is an *arbitrary* (backward) Markov kernel.

• "Proof" of validity:

$$\int \pi_2(x_2) L_1(x_2, x_1) dx_1 = \pi_2(x_2) \underbrace{\int L_1(x_2, x_1) dx_1}_{=1! \text{ whatever being } L_1} = \pi_2(x_2)$$

3.11– How to use local moves

- Similarly at time n,

$$Z_n^{-1} w_n(x_n) = \frac{\pi_n(x_n)}{\mu_n(x_n)} \text{ IMPOSSIBLE so USE } Z_n^{-1} w_n(x_{1:n}) = \frac{\widetilde{\pi}_n(x_{1:n})}{\mu_n(x_{1:n})}$$

where $\{\widetilde{\pi}_n\}$ is defined using an *sequence of arbitrary backwards* Markov kernels $\{L_n\}$

$$\text{Artificial joint target} \quad : \quad \widetilde{\pi}_n(x_{1:n}) = \pi_n(x_n) \prod_{k=1}^{n-1} L_k(x_{k+1}, x_k),$$

$$\text{Joint importance distribution} \quad : \quad \mu_n(x_{1:n}) = \mu_1(x_1) \prod_{k=2}^{n} K_k(x_{k-1}, x_k).$$

- "Proof" of validity

$$\int \widetilde{\pi}_n(x_{1:n})\, dx_{1:n-1} = \pi_n(x_n) \underbrace{\int \prod_{k=1}^{n-1} L_k(x_{k+1}, x_k)\, dx_{1:n-1}}_{=1! \text{ whatever being } \{L_k\}} = \pi_n(x_n).$$

3.12– How to select the backward Markov kernels

• **No free lunch**: By extending the integration space, the variance of the importance weights can only increase.

• The optimal kernel $\{L_{n-1}\}$ is the one bringing us back to the case where there is no space extension; i.e.

$$L_{n-1}^{\text{opt}}(x_n, x_{n-1}) = \frac{\mu_{n-1}(x_{n-1}) K_n(x_{n-1}, x_n)}{\mu_n(x_n)}$$

• The result follows straightforwardly from the forward-backward formula for Markov processes

$$\mu_n(x_{1:n}) = \mu_1(x_1) \prod_{k=2}^{n} K_k(x_{k-1}, x_k) = \mu_n(x_n) \prod_{k=2}^{n} L_{k-1}^{\text{opt}}(x_k, x_{k-1})$$

• L_{n-1}^{opt} cannot typically be computed (though there are important exceptions) but can be properly approximated in numerous cases (see later). *Even if an approximation is used, the estimates are still asymptotically consistent.*

3.13– Connections to standard SMC methods

• We are back to "standard" SMC methods where one is interested in sampling from a sequence of (artificial) distributions $\{\widetilde{\pi}_n\}$ whose dimension is increasing over time.

• **Key difference**: Given $\{K_n\}$, $\{\widetilde{\pi}_n\}$ has been constructed in a "clever" way such that

$$\int \widetilde{\pi}_n \left(x_{1:n}\right) dx_{1:n-1} = \pi_n \left(x_n\right)$$

whereas usually the sequence of targets $\{\widetilde{\pi}_n\}$ is fixed
and $\{K_n\}$ is designed accordingly.

• Because we cannot use $\{L_n^{\text{opt}}\}$ at each time step, the variance of the weights typically increases over time and it is necessary to resample.

Initialization; $n = 1$.

For $i = 1, ..., N$, sample $X_1^{(i)} \sim \mu_1(\cdot)$ and set

$$W_1^{(i)} \propto \frac{\pi_1\left(X_1^{(i)}\right)}{\mu_1\left(X_1^{(i)}\right)}.$$

Resample $\left\{W_1^{(i)}, X_1^{(i)}\right\}$ to obtain N new particles $\left\{N^{-1}, X_1^{(i)}\right\}$.

At time n; $n > 1$.

For $i = 1, ..., N$, sample $X_n^{(i)} \sim K_n\left(X_{n-1}^{(i)}, \cdot\right)$ and set

$$W_n^{(i)} \propto \frac{\widetilde{\pi}_n\left(X_{1:n}^{(i)}\right)}{\mu_n\left(X_{1:n}^{(i)}\right)} \propto W_{n-1}^{(i)} \frac{\pi_n\left(X_n^{(i)}\right) L_{n-1}\left(X_n^{(i)}, X_{n-1}^{(i)}\right)}{\pi_{n-1}\left(X_{n-1}^{(i)}\right) K_n\left(X_{n-1}^{(i)}, X_n^{(i)}\right)}.$$

Resample $\left\{W_n^{(i)}, X_n^{(i)}\right\}$ to obtain N new particles $\left\{N^{-1}, X_n^{(i)}\right\}$.

- Monte Carlo approximation

$$\widehat{\pi}_n(dx) = \sum_{i=1}^{N} W_n^{(i)} \delta_{X_n^{(i)}}(dx).$$

- Ratio of normalizing constants

$$\begin{aligned}
\frac{Z_n}{Z_{n-1}} &= \frac{\int \gamma_n(x_n)\, dx_n}{\int \gamma_{n-1}(x_{n-1})\, dx_{n-1}} \\
&= \int \frac{\gamma_n(x_n)\, L_{n-1}(x_n, x_{n-1})}{\gamma_{n-1}(x_{n-1})\, K_n(x_{n-1}, x_n)} \pi_{n-1}(dx_{n-1})\, K_n(x_{n-1}, dx_n) \\
\Rightarrow \quad \widehat{\frac{Z_n}{Z_{n-1}}} &= \sum_{i=1}^{N} W_{n-1}^{(i)} \frac{\gamma_n\left(X_n^{(i)}\right) L_{n-1}\left(X_n^{(i)}, X_{n-1}^{(i)}\right)}{\gamma_{n-1}\left(X_{n-1}^{(i)}\right) K_n\left(X_{n-1}^{(i)}, X_n^{(i)}\right)}.
\end{aligned}$$

- Like in MCMC, in practice one typically wants to use a mixture of moves

$$K_n(x, x') = \sum_{m=1}^{M} \alpha_{n,m}(x) K_{n,m}(x, x')$$

where $\alpha_{n,m}(x) > 0$, $\sum_{m=1}^{M} \alpha_{n,m}(x) = 1$ and $\{K_{n,m}\}$ is a collection of transition kernels.

- Importance weight can be computed using standard formula but can be too computationally intensive if M is large.

- L_{n-1}^{opt} can be difficult to approximate if M is large.

3.16– Extensions

• Alternative importance sampling on joint space (e.g. Auxiliary Particle Filters by Pitt & Shephard) by introducing explicitly a discrete latent variable M_n

$$\Pr\left(M_n = m\right) = \alpha_{n,m}\left(x\right)$$

and performing importance sampling on the extended space.

• The resulting incremental importance weight becomes

$$\frac{\pi_n\left(x'\right)\beta_{n-1,m}\left(x'\right)L_{n-1,m}\left(x',x\right)}{\pi_{n-1}\left(x\right)\alpha_{n,m}\left(x\right)K_{n,m}\left(x,x'\right)} \text{ instead of } \frac{\pi_n\left(x'\right)L_{n-1}\left(x',x\right)}{\pi_{n-1}\left(x\right)K_n\left(x,x'\right)}$$

where $L_{n-1}\left(x',x\right)$ is the artificial backward Markov kernel

$$L_{n-1}\left(x',x\right) = \sum_{m=1}^{M}\beta_{n-1,m}\left(x'\right)L_{n-1,m}\left(x',x\right)$$

• Optimal choice for $\{\beta_{n-1,m}, L_{n-1,m}\}$ follows straightforwardly.

3.17– Convergence Results - Central Limit Theorem

• Convergence results follow from general results on Feynman-Kac formula (see Del Moral, 2004).

• When no resampling is performed, one has

$$\sqrt{N}\left(E_{\widehat{\pi}_n}\left[\varphi\right]-E_{\pi_n}\left[\varphi\right]\right)\Rightarrow\mathcal{N}\left(0,\int\frac{\widetilde{\pi}_n^2(x_{1:n})}{\mu_n(x_{1:n})}\left(\varphi\left(x_n\right)-E_{\pi_n}\left(\varphi\right)\right)^2dx_{1:n}\right)$$

When multinomial resampling is used at each iteration, one has

$$\sqrt{N}\left(E_{\widehat{\pi}_n}\left[\varphi\right]-E_{\pi_n}\left[\varphi\right]\right)\Rightarrow\mathcal{N}\left(0,\sigma_{SMC,n}^2\left(\varphi\right)\right),$$

$$\sigma_{SMC,n}^2\left(\varphi\right)=\int\frac{\widetilde{\pi}_n^2(x_1)}{\mu_1(x_1)}\left(\int\varphi\left(x_n\right)\widetilde{\pi}_n\left(x_n\middle|x_1\right)dx_n-E_{\pi_n}\left(\varphi\right)\right)^2dx_1$$

$$+\sum_{k=2}^{n-1}\int\frac{\left(\widetilde{\pi}_n(x_k)L_{k-1}(x_k,x_{k-1})\right)^2}{\pi_{k-1}(x_{k-1})K_k(x_{k-1},x_k)}\left(\int\varphi\left(x_n\right)\widetilde{\pi}_n\left(x_n\middle|x_k\right)dx_n-E_{\pi_n}\left(\varphi\right)\right)^2dx_{k-1:k}$$

$$+\int\frac{\left(\pi_n(x_n)L_{n-1}(x_n,x_{n-1})\right)^2}{\pi_{n-1}(x_{n-1})K_n(x_{n-1},x_n)}\left(\varphi\left(x_n\right)-E_{\pi_n}\left(\varphi\right)\right)^2dx_{n-1:n}.$$

• Under mixing assumptions, $\sigma_{SMC,n}\left(\varphi\right)$ upper bounded over time.

3.18– Convergence Results - Asymptotic Bias

- When no resampling is performed, one has

$$N\left(E_{\widehat{\pi}_n}\left[\varphi\right] - E_{\pi_n}\left[\varphi\right]\right) \rightarrow -\int \frac{\pi_n^2(x_{1:n})}{\mu_n(x_{1:n})}\left(\varphi\left(x_n\right) - E_{\pi_n}\left(\varphi\right)\right) dx_{1:n}$$

- When multinomial resampling is used at each iteration (Del Moral, D. & Peters, 2004), one has

$$N\left(E_{\widehat{\pi}_n}\left[\varphi\right] - E_{\pi_n}\left[\varphi\right]\right) \rightarrow b_{SMC,n}\left(\varphi\right),$$

where with multinomial resampling

$$b_{SMC,n}\left(\varphi\right) = -\int \frac{\widetilde{\pi}_n^2(x_1)}{\mu_1(x_1)}\left(\int \varphi\left(x_n\right)\widetilde{\pi}_n\left(x_n\middle|\, x_1\right) dx_n - E_{\pi_n}\left(\varphi\right)\right) dx_1$$

$$-\sum_{k=2}^{n-1}\int \frac{\left(\widetilde{\pi}_n(x_k)L_{k-1}(x_k,x_{k-1})\right)^2}{\pi_{k-1}(x_{k-1})K_k(x_{k-1},x_k)}\left(\int \varphi\left(x_n\right)\widetilde{\pi}_n\left(x_n\middle|\, x_k\right) dx_n - E_{\pi_n}\left(\varphi\right)\right) dx_{k-1:k}$$

$$-\int \frac{\left(\pi_n(x_n)L_{n-1}(x_n,x_{n-1})\right)^2}{\pi_{n-1}(x_{n-1})K_n(x_{n-1},x_n)}\left(\varphi\left(x_n\right) - E_{\pi_n}\left(\varphi\right)\right) dx_{n-1:n}.$$

- Under mixing assumptions, $\left|b_{SMC,n}\left(\varphi\right)\right|$ upper bounded over time.

4.1– How to design an SMC method for sampling from fixed target

• **First step**: Build a sequence of distributions $\{\pi_n\}$ going from π_1 easy to sample/approximate to $\pi_P = \pi$; e.g. $\pi(x) \propto [\mu_1(x)]^{\eta_n} [\pi(x)]^{1-\eta_n}$ where μ_1 easy to sample and $\eta_1 = 1$, $\eta_n < \eta_{n-1}$ with $\eta_P = 0$.
• **Second step**: Introduce a sequence of transition kernels $\{K_n\}$; e.g. K_n MCMC sampler of invariant distribution π_n.
• **Third step**: Introduce a sequence of backward kernels $\{L_n\}$ equal/approximating L_n^{opt}; e.g.

$$L_{n-1}(x_n, x_{n-1}) = \frac{\pi_{n-1}(x_{n-1}) K_n(x_{n-1}, x_n)}{\int \pi_{n-1}(x_{n-1}) K_n(x_{n-1}, x_n) dx_{n-1}}$$

$$\Rightarrow \quad \alpha_n(x_{n-1}, x_n) = \frac{\pi_n(x_n)}{\int \pi_{n-1}(x_{n-1}) K_n(x_{n-1}, x_n) dx_{n-1}}$$

$$L_{n-1}(x_n, x_{n-1}) = \frac{\pi_n(x_{n-1}) K_n(x_{n-1}, x_n)}{\pi_n(x_n)} \Rightarrow \alpha_n(x_{n-1}, x_n) = \frac{\pi_n(x_{n-1})}{\pi_{n-1}(x_{n-1})}$$

4.2– Bayesian Analysis of Finite Mixture of Gaussians

- Model

$$Y_i | \, X_i \sim \mathcal{N}\left(\mu_{X_i}, \sigma^2_{X_i}\right),$$

$$\Pr\left(X_i = k\right) = \pi_k \text{ where } k = 1, ..., M$$

- Standard conjugate priors on $\theta = \left(\pi_k, \mu_k, \sigma^2_k\right)$, no identifiability constraint, posterior is a mixture of $M!$ components.

- Simulations with $M = 4$, components "far" from each other.

- Gibbs sampler to sample from $p\left(\theta, x_{1:T} | \, y_{1:T}\right)$ get stuck in one mode.

• To sample $p(\theta, x_{1:T} | y_{1:T})$, set $\pi_n(\theta, x_{1:T}) \propto p(\theta, x_{1:T}) p(y_{1:T} | \theta, x_{1:T})^{\eta_n}$ where $n \in \{1, \ldots, P\}$, $N = 1000$.

• For moderate P, SMC discovers the 4! modes and provide good estimates of $\mathbb{E}[\mu_i | Y_{1:T}]$.

• Generally, resampling helps when number of intermediate distributions P is low, otherwise no significant difference.

• Use SMC as exploratory techniques?

5.1– Sequential Bayesian Trans-dimensional Estimation

• At time t, time occurrences assumed to follow an inhomogeneous Poisson process of intensity $\lambda : R^+ \rightarrow R^+$

$$p_t\left(y_{1:l_t} | \{\lambda(u)\}_{u \leq t}\right) = \exp\left(-\int_0^t \lambda(u)\, du\right) \prod_{l=1}^{l_t} \lambda(y_l).$$

• We want to estimate unknown intensity $\lambda(t)$ sequentially in time.

• Simple piecewise constant model for $\lambda(t)$

$$\lambda(t) = \sum_{m=1}^{k} \lambda_m 1_{[\tau_{m-1}, \tau_m)}(t)$$

• The number of steps k, their amplitudes $\lambda_{1:k+1}$ and the knot points $\tau_{1:k}$ are assumed unknown $\Rightarrow$ Set following time-dependent prior distribution

$$p_t(k, \lambda_{1:k+1}, \tau_{1:k}) = p_t(k)\, p(\lambda_{1:k+1} | k)\, p_t(\tau_{1:k} | k)$$

where $p_t(k)$ Poisson $\lambda_q t$, $p_t(\tau_{1:k} | k)$ uniform order statistics on $[0, t]$ and $\lambda_1 \sim \mathcal{G}(\alpha, \beta)$ and $\lambda_l | \lambda_{l-1} \sim \mathcal{G}\left(\lambda_{l-1}^2/\chi; \lambda_{l-1}/\chi\right)$.

5.1– Sequential Bayesian Trans-dimensional Estimation

- Sequential estimation of posterior distributions over times $n\Delta T$

$$\pi_n\left(k, \lambda_{1:k+1}, \tau_{1:k}\right) = p_{n\Delta T}\left(k, \lambda_{1:k+1}, \tau_{1:k} \middle| y_{1:l_{n\Delta T}}\right)$$

where ΔT is a time interval defined by the user.

- These distributions are defined on $E = \cup_{k=0}^{\infty} \{k\} \times \vartheta_k$ where $\vartheta_k = \left\{\tau_{1:k} \in \left(R^+\right)^k ; 0 < \tau_1 < \ldots < \tau_k\right\} \times \left(R^+\right)^{k+1}$, the support of π_n being reduced to $\left\{\tau_{1:k} \in \left(R^+\right)^k ; 0 < \tau_1 < \ldots < \tau_k < n\Delta T\right\} \times \left(R^+\right)^{k+1}$.

- Combinations of birth/death and extend steps.

5.2– Coal mining data