SMC 2017 $\boxed{1}$

Terminology

Assume y data, θ parameters

Full probabilistic model: $p(\theta, y) = \underbrace{p(y|\theta)}_{\substack{\text{data distribution} \\ \text{likelihood}}} \underbrace{p(\theta)}_{\text{prior}}$

Likelihood function $\mathcal{L}(\theta; y) = p(Y = y | \theta)$ function of unknown parameters, not a distribution anymore

Bayesian approach: $\underbrace{p(\theta|y)}_{\text{posterior distribution}} = \frac{p(\theta, y)}{p(y)} = \frac{p(y|\theta)\, p(\theta)}{\underbrace{\int p(y|\theta)\, p(\theta)\, d\theta}_{\text{normalization constant / model evidence}}}$

Prediction: $p(\bar{y}|y) = \int p(\bar{y}|\theta)\, p(\theta|y)\, d\theta$

possible, uncertain future value

conditional prob.

$\left(\int p(\bar{y}, \theta | y)\, d\theta\right)$ marginalisation

Linear Gaussian state space model (LG-SSM)

simple autoreg. model

$X_t = A X_{t-1} + v_t \quad , \quad v_t \sim \mathcal{N}(0, Q)$ state model

process noise

X_{t-1}: state at time t

$Y_t = C X_t + E_t \quad , \quad E_t \sim \mathcal{N}(0, R)$ measurement model

Y_t: measurement at time t

E_t: measurement noise

Transition model (dynamics):

$$p(x_t | x_{t-1}) = \mathcal{N}(x_t | A x_{t-1}, Q)$$

Measurement model:

$$p(y_t | x_t) = \mathcal{N}(y_t | C x_t, R)$$

model distribution

Importance sampling

Can be used to solved integrals of the type

$$I(f) = \mathbb{E}_\pi(f(X)) = \int f(x)\,\pi(x)\,dx \quad (\text{where } X \sim \pi(x))$$

without requiring exact samples from the target distribution.

Note that

$$E_\pi(f(X)) = \int f(x)\,\pi(x)\,dx = \int f(x)\,\frac{\pi(x)}{q(x)}\,q(x)\,dx$$

proposal distribution

as long as division by zero is avoided

Set: $\omega(x) := \frac{\pi(x)}{q(x)}$, weight function, then

$$E_\pi(f(X)) = \int f(x)\,w(x)\,q(x)\,dx = E_q((f\cdot\omega)(X)) \quad (\text{here } X \sim q(x))$$

Results in two steps:

1. Sample from the proposal distribution
$$x^i \sim q(x),\ i = 1,\dots,N$$
2. Compute weights
$$w^i = \omega(x^i) = \frac{\pi(x^i)}{q(x^i)},\quad i = 1,\dots,N$$

Then $\hat{I}^N(f) = \frac{1}{N}\sum_{i=1}^{N} w^i f(x^i)$

Problems: a) How to choose q?, b) How to calculate w^i?

a) Experience ...

b) If $\pi(x) = \frac{\tilde{\pi}(x)}{Z}$ can be calculated except for a normalization constant, then $Z = \int \tilde{\pi}(x)\,dx = \int \frac{\tilde{\pi}(x)}{q(x)}\,q(x)\,dx = \int \omega(x)\,q(x)\,dx$ which can be approximated by Monte Carlo as well.

Nonlinear filtering problem

What if we do not know $\pi(x)$ up to a normalisation constant?

Assume that at time $t-1$ we have

$$\hat{p}^N(x_{t-1} \mid y_{1:t-1}) = \sum_{i=1}^{N} w_{t-1}^i \, \delta_{x_{t-1}^i}(x_{t-1}) \qquad (*)$$

Insert $(*)$ into the time update integral, then

$$\hat{p}^N(x_t \mid y_{1:t-1}) = \int p(x_t \mid x_{t-1}) \sum_{i=1}^{N} w_{t-1}^i \, \delta_{x_{t-1}^i}(x_{t-1}) \, dx_{t-1} =$$

$$= \sum_{i=1}^{N} w_{t-1}^i \, p(x_t \mid X_{t-1} = x_{t-1}^i)$$

The Measurement update then is

$$p(x_t \mid y_{1:t}) \approx \frac{p(y_t \mid x_t)}{p(y_t \mid y_{1:t-1})} \sum_{i=1}^{N} w_{t-1}^i \, p(x_t \mid X_{t-1} = x_{t-1}^i) \qquad \boxed{\Pi}$$

This approx. of the filter pdf can be evaluated up to normalisation, and it can thus be targeted by an importance sampler at time t.

Guided by $\boxed{\Pi}$ we choose a mixture distribution as proposal:

$$q(x_t \mid y_{1:t}) = \sum_{i=1}^{N} v_{t-1}^i \, q(x_t \mid x_{t-1}^i, y_t)$$

design choices

Markov property, older values are not necessary

We can now compute the weights as

$$\tilde{\omega}_t^i = \frac{\bar{\pi}(x_t^i)}{q(x_t^i)} = \frac{p(y_t \mid x_t^i) \sum_{j=1}^{N} w_{t-1}^j \, p(x_t^i \mid x_{t-1}^j)}{\sum_{j=1}^{N} v_{t-1}^j \, q(x_t^i \mid x_{t-1}^j, y_t)}$$

→ we obtain a new set of weighted particles $\{x_t^i, w_t^i\}_{i=1}^N$ approx. $p(x_t \mid y_{1:t})$

<u>Problem:</u> $O(N^2)$ computational complexity

SMC 2017

$\boxed{4}$

Pragmatic solution:

$$q(x_t|y_{1:t}) = \hat{p}^N(x_t|y_{1:t-1}) = \sum_{j=1}^{N} w_{t-1}^j \, p(x_t|x_{t-1}^j)$$

then $\tilde{\omega}_t^i = p(y_t|x_t^i)$

<u>Central limit theorem for Importance Sampling:</u>

Imp. Samp. estimate of $I(f) = \int f(x)\pi(x)dx$ is

$$\hat{I}_N^{IS}(f) = \sum_{i=1}^{N} \frac{w(x^i)}{\sum_{j=1}^{N} w(x^j)} f(x^i) = \frac{\frac{1}{N}\sum_{i=1}^{N} w(x^i) f(x^i)}{\frac{1}{N}\sum_{i=1}^{N} w(x^i)}$$

Let $g(x) = w(x)f(x)$ and $\bar{g} = \frac{1}{N}\sum_{i=1}^{N} w(x^i)f(x^i)$, $\bar{\omega} := \frac{1}{N}\sum_{i=1}^{N} w(x^i)$

(sample means), then $\hat{I}_N^{IS}(f) = \frac{\bar{g}}{\bar{\omega}}$

$\bar{g}, \bar{\omega}$ are both std. MC estimates, therefore SLLN & CLT apply, i.e.

$$E(\bar{g}) = E_q(g(X)) = \int g(x)\,q(x)\,dx = \int \frac{\pi(x)}{q(x)} f(x) q(x)\,dx = I(f)$$

$$E(\bar{\omega}) = E_q(w(X)) = \int \frac{\pi(x)}{q(x)} q(x)\,dx = 1$$

Consider a Taylor expansion of $\frac{1}{\bar{\omega}}$ around its mean of 1 (= delta method)

$$\hat{I}_N^{IS}(f) = \frac{\bar{g}}{\bar{\omega}} = \bar{g}\left(1 - (\bar{\omega}-1) + (\bar{\omega}-1)^2 + \ldots\right) =$$

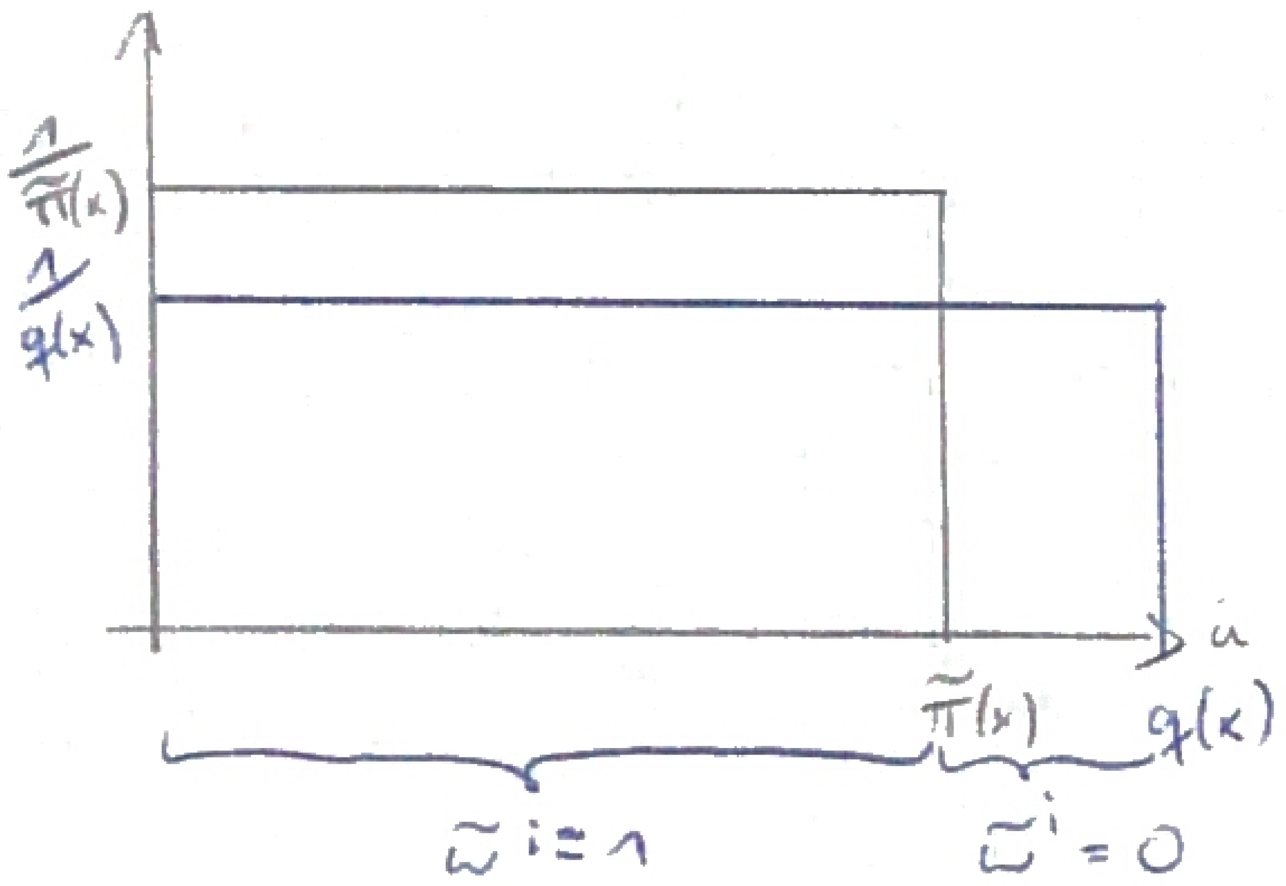
$$= \underbrace{(\bar{g} - I(f))}_{\text{both terms will be small for large } N}[1 - \underbrace{(\bar{\omega}-1)} + \ldots] + I(f)[1 - (\bar{\omega}-1) + \ldots] =$$

$$= I(f) + (\bar{g} - I(f)) - I(f)(\bar{\omega}-1) - (\bar{g} - I(f))(\bar{\omega}-1) +$$

$$+ I(f)(\bar{\omega}-1)^2 + \text{higher order terms}$$

then $E(\hat{I}_N^{IS}(f)) = I(f) - \frac{Cov_q(g(X), w(X))}{N} + I(f)\frac{Var_q(w(X))}{N} + O\left(\frac{1}{N^2}\right)$

Importance sampling with an auxiliary variable



rejection sampling

Particle filtering with auxiliary variables

Unnormalized target pdf of the particle filter at time t

$$p(y_t|x_t) \cdot \underbrace{\sum_{i=1}^{N} w_{t-1}^{i}\, p(x_t|x_{t-1}^{i})}_{\text{proposal used in bootstrap particle filter}}$$

Introduce an auxiliary variable A_t representing the mixture index.
The joint target for (X_t, A_t) is proportional to

$$p(y_t|x_t)\, w_{t-1}^{a_t}\, p(x_t|x_{t-1}^{a_t}) \qquad (*)$$

N.B. Marginalizing over the auxiliary variable we get

$$\sum_{a_t=1}^{N} p(y_t|x_t^{a_t})\, w_{t-1}^{a_t}\, p(x_t|x_{t-1}^{a_t})$$

as required.

We can thus sample pairs $\{(x_t^i, a_t^i)\}_{i=1}^N$ from $(*)$ (using imp. samp.) and use the x's (with weights) to approximate the filtering distribution at time t. Specifically, select proposal for (X_t, A_t) as

$$q(x_t, a_t | y_{1:t}) = \nu_{t-1}^{a_t}\, q(x_t | x_{t-1}^{a_t}, y_t) \qquad \boxed{}$$

Thus,

- A_t has a categorical distribution on $\{1, \ldots, N\}$ with probabilities $\{\nu_{t-1}^i\}_{i=1}^N$

- The conditional distribution for x_t given $A_t = a_t$ is $q(x_t \mid x_{t-1}^{a_t}, y_t)$

The importance weights are given by the ratio of the joint target (A) and the joint proposal (B), i.e.

$$\frac{w_{t-1}^{a_t}}{v_{t-1}^{a_t}} \frac{p(y_t \mid x_t)\, p(x_t \mid x_{t-1}^{a_t})}{q(x_t \mid x_{t-1}^{a_t}, y_t)}$$

No sum!

The joint proposal for (X_t, A_t) is locally optimal if it exactly matches the target for (X_t, A_t) at time t. Then $w_t^i = \frac{1}{N}$

Joint target $\propto$ $w_{t-1}^{a_t}\, p(y_t \mid x_t)\, p(x_t \mid x_{t-1}^{a_t})$

Joint proposal $=$ $v_{t-1}^{a_t}\, q(x_t \mid x_{t-1}^{a_t}, y_t)$

To match proposal = target, we first need to normalize the target.

Note that

$$p(y_t \mid x_t)\, p(x_t \mid x_{t-1}) = p(y_t, x_t \mid x_{t-1}) = p(x_t \mid x_{t-1}, y_t)\, p(y_t \mid x_{t-1})$$

Thus, the normalized joint target can be written as

$$\underbrace{\frac{w_{t-1}^{a_t}\, p(y_t \mid x_{t-1}^{a_t})}{\sum_{j=1}^{N} w_{t-1}^{j}\, p(y_t \mid x_{t-1}^{j})}}_{\text{marginal distribution of the auxiliary variable } A_t} \underbrace{p(x_t \mid x_{t-1}^{a_t}, y_t)}_{\text{conditional distribution of } X_t \mid (A_t = a_t)}$$

Using

- $v_{t-1}^i \propto w_{t-1}^i\, p(y_t \mid x_{t-1}^i)$, $i = 1, \ldots, N$
- $q(x_t \mid x_{t-1}, y_t) = p(x_t \mid x_{t-1}, y_t)$

we are effectively sampling from the joint target for (X_t, A_t).

Therefore, $\tilde{w}_t^i = \text{const}$ and $w_t^i = \frac{1}{N}$.

This is referred to the fully adapted particle filter.

SMC 2017

4

The locally optimal proposals, based on $p(x_t \mid x_{t-1}, y_t)$ and $p(y_t \mid x_{t-1})$, can be computed when $y_t \mid x_t$ is conjugate to $x_t \mid x_{t-1}$.

Example: Gaussian model with nonlinear dynamics

$$\begin{cases} X_t = f(X_{t-1}) + V_t, & V_t \sim \mathcal{N}(0, Q) \\ Y_t = C X_t + E_t, & E_t \sim \mathcal{N}(0, R) \end{cases}$$

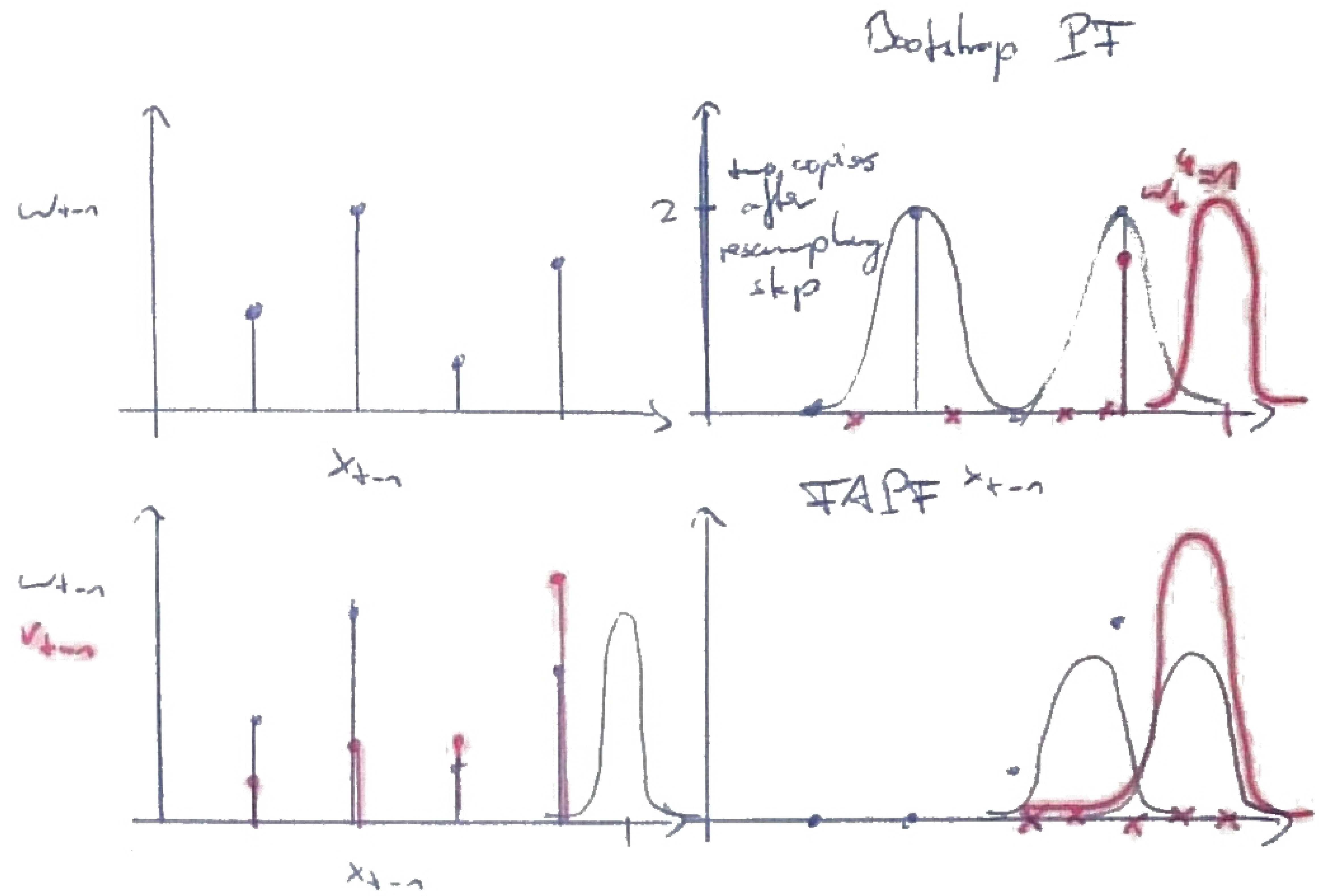
$$\Rightarrow \begin{cases} p(x_t \mid x_{t-1}) = \mathcal{N}(x_t;\, f(x_{t-1}), Q) \\ p(y_t \mid x_t) = \mathcal{N}(y_t;\, C x_t, R) \end{cases}$$

conjugate relationship given x_{t-1}

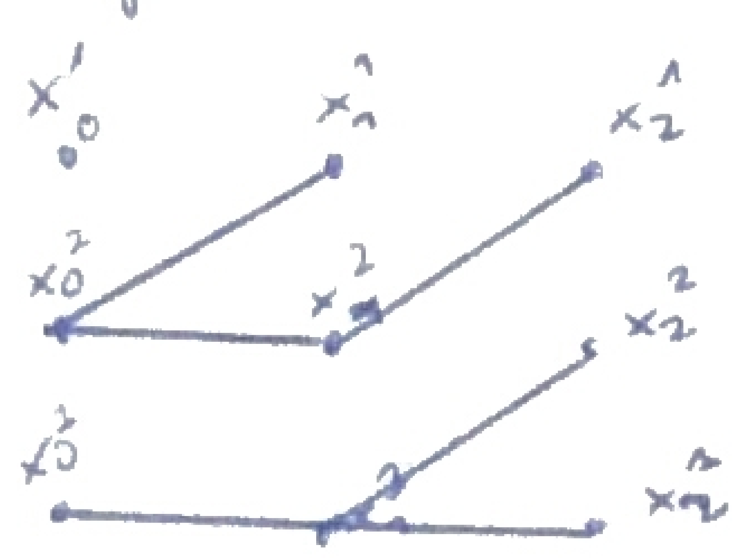
Example (Illustration)

$X_t = X_{t-1} + V_t$

$Y_t = X_t + E_t$



Example (ancestor indices)



$\underline{t=1}$: $a_1^1 = 2$, $a_1^2 = 2$, $a_1^3 = 3$

$\underline{t=2}$: $a_2^1 = 2$, $a_2^2 = 3$, $a_2^3 = 3$

Ancestral path for particle x_2^1

$$x_{0:2}^1 = \{ x_{0:1}^{a_2^1}, x_2^1 \} = \{ x_0^{a_1^2}, x_1^2, x_2^1 \} = \{ x_0^2, x_1^2, x_2^1 \}$$

Likelihood computation

$$p(y_{1:T}\mid\theta) = \prod_{t=1}^{T} p(y_t\mid y_{1:t-1},\theta)$$

with $p(y_t\mid y_{1:t-1},\theta) = \int p(y_t,x_t\mid y_{1:t-1},\theta)\,dx_t =$

$$= \int p(y_t\mid x_t,\theta)\,\underbrace{p(x_t\mid y_{1:t-1},\theta)}\,dx_t =$$

Markov model: $y_{1:t-1}$ does not contain any information that isn't in x_t

(bootstrap PF) $\approx \sum_{i=1}^{N}\frac{1}{N}\delta_{x_t^i}(x_t)$

$$= \cancel{\int p}\ \frac{1}{N}\sum_{i=1}^{N} p(y_t\mid x_t^i,\theta) = \frac{1}{N}\sum_{i=1}^{N}\tilde{\omega}_t^i$$

Together: $p(y_{1:T}\mid\theta) \approx \prod_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{\omega}_t^i\right)$ Could be numerically problematic!

Solution: Use shifted log-weights

Define $v_t^i = \log\tilde{\omega}_t^i - C_t$, $i=1,\dots,N$

where $C_t := \max\{\log\tilde{\omega}_t^1,\dots,\log\tilde{\omega}_t^N\}$, then $\tilde{\omega}_t^i = \exp(v_t^i + C_t)$

$$\log p(y_{1:T}\mid\theta) = \underbrace{\log p(y_T\mid y_{1:T-1},\theta)} + \log p(y_{1:T-1}\mid\theta)$$

$$\approx \log\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{\omega}_t^i\right) = c_t + \log\left(\sum_{i=1}^{N} e^{v_t^i}\right) - \log N$$

Properties of the likelihood estimator

A particle filter that is run for time steps $t=0,\dots,T$ samples the random variables

$$\mathbf{X}_t = \{x_t^i\}_{i=1}^{N}, \quad t=0,\dots,T$$

$$\mathbf{A}_t = \{A_t^i\}_{i=1}^{N}, \quad t=1,\dots,T$$

The distributions from which these are sampled are, for the bootstrap PF:

Initialisation: $\mathbf{X}_0 \sim \prod_{i=1}^{N} p(x_0^i)$

SMC 2019

Resampling. $A_t \mid (X_{t-1} = x_{t-1}) \sim$

$$\sim \prod_{i=1}^{N} C(a_t^i \mid \{w_{t-1}^j\}_{j=1}^N) = \prod_{i=1}^{N} w_{t-1}^{a_t^i}$$

Propagation. $X_t \mid (X_{t-1} = x_{t-1}, A_t = a_t) \sim \prod_{i=1}^{N} p(x_t^i \mid x_{t-1}^{a_t^i})$

<u>Unbiasedness of the likelihood estimator</u>

Proof idea. Solve the integral

$$\mathbb{E}_{\psi_{N,T}}[\hat{Z}] = \int \prod_{t=1}^{T} \Big\{ \frac{1}{N} \sum_{i=1}^{N} \tilde{\omega}_t^i \Big\} \psi_{N,T}(x_{0:T}, a_{1:T}) \, d x_{0:T} \, d a_{0:T} =$$

$\tilde{\omega}_t^i = p(y_t \mid x_t^i)$

mix of continuous and discrete variables

$$= \int \prod_{t=1}^{T-1} \Big\{ \frac{1}{N} \sum_{i=1}^{N} \tilde{\omega}_t^i \Big\} \psi_{N,T-1}(x_{0:T-1}, a_{1:T-1}) \times$$

$$\times \underbrace{\Big[\int \Big\{ \frac{1}{N} \sum_{i=1}^{N} \tilde{\omega}_T^i \Big\} \prod_{i=1}^{N} w_{T-1}^{a_T^i} \, p(x_T^i \mid x_{T-1}^{a_T^i}) \, dx_T \, da_T \Big]}_{\text{Integral over } (x_T, a_T) \text{ can be computed}} \, dx_{0:T-1} \, da_{1:T-1}$$

$$= \dots = \sum_{i=1}^{N} w_{T-1}^i \, p(y_T \mid x_{T-1}^i)$$

Proceeding similarly, integrating over (x_{T-1}, a_{T-1}), etc. gives the result.

Particle filter is a smart set of choosing and combining random numbers:

1) In the prior
2) In the resampling step
3) In the propagation step

Markov Chain Monte Carlo

We want Markov chains $\{x[m]\}_{m=1}^{M}$, so that for $M \to \infty$

$$\frac{1}{M} \sum_{m=1}^{M} f(x[m]) \xrightarrow{a.s.} \int f(x)\, \pi(x)\, dx$$

Such Markov chains can be constructed by

1) Sample a candidate x' from a proposal distribution

$$x' \sim q(x \mid x[m])$$

2) Choose the candidate sample x' as the next state with probability

acceptance probability

$$\alpha = \min\left(1, \frac{\pi(x')}{\pi(x[m])} \frac{q(x[m] \mid x')}{q(x' \mid x[m])}\right)$$

i.e. $$x[m+1] = \begin{cases} x' & \text{with prob. } \alpha \\ x[m] & \text{with prob. } 1-\alpha \end{cases}$$

This is the Metropolis-Hastings algorithm.

To use MH we need to: (in case of parameter estimation inference)

1) Decide on a proposal q
2) Compute the acceptance probability α

$$\alpha = \frac{p(\theta' \mid y_{1:T})}{p(\theta[m] \mid y_{1:T})} \cdot \frac{q(\theta[m] \mid \theta')}{q(\theta' \mid \theta[m])} =$$

$$= \left[\frac{p(y_{1:T} \mid \theta')\, p(\theta')}{\cancel{p(y_{1:T})}} \Big/ \frac{p(y_{1:T} \mid \theta[m])\, p(\theta[m])}{\cancel{p(y_{1:T})}} \right] \cdot \frac{q(\cdot)}{q(\cdot)}$$

Particle Metropolis Hastings (PMH)

The particle filter provides:

- non-negative and unbiased

estimate $\hat{z}$ of the likelihood $p(y_{1:T} \mid \theta)$. The estimate $\hat{z}$ is itself a random variable $\hat{z} \sim \psi(z \mid \theta, y_{1:T})$.

Consider an extended model, where $\hat{z}$ is included as an auxiliary variable

$$(\theta, \hat{z}) \sim \psi(\theta, z \mid y_{1:T}) = \psi(z \mid \theta, y_{1:T}) \cdot p(\theta \mid y_{1:T}) =$$
$$= \frac{p(y_{1:T} \mid \theta)\, p(\theta)}{p(y_{1:T})} \cdot \psi(z \mid \theta, y_{1:T})$$

Key trick: Define a new joint target distribution over $(\theta, \hat{z})$ by simply replacing $p(y_{1:T} \mid \theta)$ with $\hat{p}(y_{1:T} \mid \theta)$ (its estimator $\hat{z}$), i.e.

$$\pi(\theta, z \mid y_{1:T}) = \frac{z \cdot p(\theta)}{p(y_{1:T})} \psi(z \mid \theta, y_{1:T})$$

Requirements on π:

1. Non-negative $\rightsquigarrow$ follows from non-negativity of $\hat{z}$
2. Normalized, $\rightsquigarrow$ follows from below
3. $\int \pi(z, \theta \mid y_{1:T})\, dz = p(\theta \mid y_{1:T})$

$$\int \pi(z, \theta \mid y_{1:T})\, dz = \frac{p(\theta)}{p(y_{1:T})} \underbrace{\int z \psi(z \mid \theta, y_{1:T})\, dz}_{= p(y_{1:T} \mid \theta) \text{ due to the unbiasedness of the PF}} = p(\theta \mid y_{1:T})$$

Gibbs sampling for dynamical systems (state-space models)

Bayesian model with parameter $\theta \sim p(\theta)$

Task: Compute the posterior $p(\theta \mid y_{1:T})$ based on a batch of observations $y_{1:T}$

Introduce the unknown states as auxiliary variables ("data augmentation"), i.e. target $p(\theta, x_{0:T} | y_{1:T})$

Gibbs sampler: Iterate, 1. $\theta^* \sim p(\theta | x_{0:T}, y_{1:T})$ ← typically much easier than sampling from $p(\theta | y_{1:T})$

2. $x_{0:T}^* \sim p(x_{0:T} | \theta^*, y_{1:T})$ ← state inference problem

N.B. If the prior $p(\theta)$ is conjugate to the "complete data likelihood" $p(x_{0:T}, y_{1:T} | \theta)$, then step 1 is analytically tractable.

If the model is linear & Gaussian, step 2 is also analytically tractable.

Conditional importance sampling

Thm $x_N(x, x^*)$ defined by the conditional importance sampling ~~procedure~~ procedure has $\pi(x)$ as stationary distribution for any $N \geq 1$.

Proof: We will prove the thm for the following equiv. formulation of the procedure:

Step 1:

- Draw $b \sim \mathcal{U}(\{1, \ldots, N\})$
- Draw $x^i \sim q(x)$, $i \neq b$
- Set $x^b = x$ ← input

Step 2:

- Compute $\tilde{\omega}^i = \frac{\tilde{\pi}(x^i)}{q(x^i)}$, $i = 1, \ldots, N$ and normalize
- Draw $b^* \sim \mathcal{C}(\{\omega^i\}_{i=1}^N)$
- Output: $x^* = x^{b^*}$

We will treat $(b, \{x^i\}_{i \neq b})$ as auxiliary variables.

Joint target: (definition)

$$\pi(x^{1:N}, b) = \frac{1}{N} \pi(x^b) \prod_{i \neq b} q(x^i)$$

<u>Fact 1:</u> The distribution of x^b under $\pi(x^{1:N}, b)$ is $\pi(x)$

<u>Fact 2:</u> If the input x is distributed according to $\pi(x)$ then after completing Step 1 we have $(x^{1:N}, b) \sim \pi(x^{1:N}, b)$

<u>Fact 3:</u> Step 2 is a Gibbs update step for the variable b with respect to $\pi(x^{1:N}, b)$

$$b^* \sim \pi(b \mid x^{1:N})$$

<u>Proof of Fact 3:</u> $\pi(b \mid x^{1:N}) \propto \pi(x^{1:N}, b) = \frac{1}{N} \underbrace{\frac{\pi(x^b)}{q(x^b)}}_{\propto\, \tilde{\omega}^b} \underbrace{\prod_{i=1}^{N} q(x^i)}_{\text{independent of } b} \propto \tilde{\omega}^b$

Normalizing the distribution, we get

$$\pi(b \mid x^{1:N}) = \frac{\tilde{\omega}^b}{\sum_{j=1}^{N} \tilde{\omega}^j} = \omega^b$$

#

It follows from a standard Gibbs sampling argument that $(x^{1:N}, b^*) \sim \pi(x^{1:N}, b)$.

Therefore $x^* = x^{b^*} \sim \pi(x)$.

■

<u>SMC samplers</u>

- Resampling weights: $\nu_{k-1}^i = w_{k-1}^i$
- Propagation: $q_k(x_k \mid x_{1:k-1}^{a_k^i}) = K_k(x_{k-1}^{a_k^i}, x_k)$ γ_k-invariant MH-kernel
- Weighting: $w_k^i \propto \cancel{\frac{w_{k-1}^{a_k^i}}{\nu_{k-1}^{a_k^i}}} \frac{\pi_k(x_{0:k}^i)}{\pi_{k-1}(x_{0:k-1}^{a_k^i})\, q(x_k^i \mid x_{0:k-1}^{a_k^i})} =$

$$= \frac{\gamma_k(x_k^i) \prod_{j=1}^{k} \lambda(x_j^{a_k^i}, x_{j-1}^{a_k^i})}{K_k(x_{k-1}^{a_k^i}, x_k^i)\, \gamma_{k-1}(x_{k-1}^i) \prod_{j=1}^{k-1} \lambda_{j-1}(x_j^{a_k^i}, x_{j-1}^{a_k^i})} = \frac{\gamma_k(x_k^i)\, \lambda_{k-1}(x_k^i, x_{k-1}^{a_k^i})}{\gamma_{k-1}(x_{k-1}^i)\, K(x_{k-1}^{a_k^i}, x_k^i)}$$

SMC 2017

$$w_k^i \propto \frac{\gamma_k(x_k)}{\gamma_{k-1}(x_{k-1})\, K_k(x_{k-1}, x_k)} \; \frac{\gamma_k(x_{k-1}) \cdot L_k(x_{k-1}, x_k)}{\gamma_k(x_k)} = \boxed{14}$$

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