Importance Sampling for DA: particle filters (sequential Monte Carlo)

Problem set up

stochastic model

model: $x_{k+1} = f(x_k) + v_k$, $v_k \sim \mathcal{N}(0, Q)$

obs : $y_{k+1} = h(x_{k+1}) + \eta_k$, $\eta_k \sim \mathcal{N}(0, R)$

$k = 0, 1, 2, \ldots$ integer, "discrete time".

transition distribution : $p(x_{k+1} | x_k) = \mathcal{N}(f(x_k), Q)$

likelihood : $p(y_{k+1} | x_{k+1}) = \mathcal{N}(h(x_{k+1}), R)$

Goal: Estimate state at time k, given observations up to time k (Recall KF!)

→ use importance sampling for $p(x_k | y_{1:k})$

$$\hookrightarrow p(x_k | y_{1:k}) = p(y_{1:k} | x_k) \frac{p(x_k)}{p(y_{1:k})}$$

$$P(A|B)\frac{P(B)}{P(A)} = \frac{P(A,B)}{\cancel{P(B)}}\frac{\cancel{P(B)}}{P(A)} = P(B|A)$$

$$= p(y_k | x_k, y_{1:k-1}) \frac{p(y_{1:k-1} | x_k)\, p(x_k)}{p(y_k | y_{1:k-1})\, p(y_{1:k-1})}$$

$$= \frac{p(y_k | x_k)\, p(x_k | y_{1:k-1})}{p(y_k | y_{1:k-1})} \propto \underbrace{p(y_k | x_k)}_{\text{we know it}} \underbrace{p(x_k | y_{1:k-1})}_{\text{we don't know it!}}$$

⇒ We cannot apply importance sampling because we do not know the target distribution up to a constant!

Instead: Solve a harder problem

Consider: $\underbrace{p(x_{0:k}|y_{1:k})}$

up to time k

· Probability of a trajectory of the stochastic model conditioned on the data up to time k.

~~Manipulations:~~

· Marginal of this distribution is

~~$p(x_{0:k}$~~

$$\int p(x_{0:k}|y_{1:k})\,dx_0 \cdots dx_{k-1} = p(x_k|y_{1:k})$$

· to get samples of $p(x_k|y_{1:k})$

(a) sample $p(x_{0:k}|y_{1:k})$

(b) marginalize the samples by "dropping the past".

$\{x^j_{0:k}\}$ is ensemble of trajectories, $j=1\dots N_e$ distributed as $p(x_{0:k}|y_{1:k})$

$\{x^j_k\}$ is ensemble of states, $j=1\dots N_e$, distributed as $p(x_k|y_{1:k})$

proof requested by Don Rossi: See attachment (*)

Manipulations

$$p(x_{0:k}|y_{1:k}) = p(y_{1:k}|x_{0:k})\frac{p(x_{0:k})}{p(y_{1:k})}$$

$$= \frac{\underbrace{p(y_k|x_{0:k})}_{=p(y_k|x_k)}\ \overbrace{p(y_{1:k-1}|x_{0:k})}^{=p(y_{1:k-1}|x_{0:k-1})}\ p(x_0:x_k)}{p(y_{1:k})}$$

because future does not inform past.

$=p(y_k|x_k)$ due to assumptions.

$$= p(y_k|x_k)\frac{p(y_{1:k-1}|x_{1:k-1})\ \overbrace{p(x_k|x_{0:k-1})}^{=p(x_k|x_{k-1})}\ p(x_{0:k-1})}{p(y_k|y_{1:k-1})\ p(y_{1:k-1})}$$

due to assumptions

$$= p(y_k|x_k)p(x_k|x_{k-1})\ \frac{p(y_{1:k-1}|x_{0:k-1})\,p(x_{0:k-1})}{p(y_{1:k-1})}\ \frac{1}{p(y_k|y_{1:k-1})}$$

$$= p(y_k|x_k)\,p(x_k|x_{k-1})\ \frac{p(y_{1:k-1}, x_{0:k-1})\,\cancel{p(x_{0:k-1})}}{\cancel{p(x_{0:k-1})}\ p(y_{1:k-1})}\cdot\frac{1}{p(y_k|y_{1:k-1})}$$

$$\underbrace{p(x_{0:k}|y_{1:k})}_{} = \underbrace{p(x_{0:k-1}|y_{1:k-1})}_{\text{Result from previous calculation}} \; \underbrace{p(x_k|x_{k-1})}_{\text{model transition}} \; \underbrace{p(y_k|x_k)}_{\text{likelihood}} \; \underbrace{\frac{1}{p(y_k|y_{1:k})}}_{\text{Normalization factor.}}$$

Almost all PFs use proposal of a similar form:

$$\boxed{q(x_{0:k}|y_{1:k}) \propto q(x_0) \prod_{j=1}^{k} q(x_j|x_{0:j-1}, y_{1:j})}$$

Step $k=1$

proposal: $q(x_{0:k}|y_1) \propto q(x_0)\, q(x_1|x_0, y_1)$

target: $p(x_{0:1}|y_1) \propto p(x_0)\, p(x_1|x_0)\, p(y_1|x_1)$

weight: $\omega_1 \propto \underbrace{\frac{p(x_0)}{q(x_0)}}_{\omega_0} \; \frac{p(x_1|x_0)\, p(y_1|x_1)}{q(x_1|x_0, y_1)}$

Step $k=2$

proposal: $q(x_{0:2}|y_{1,2}) \propto q(x_0)\, q(x_1|x_0, y_1)\, q(x_2|x_1, y_2)$

target : $p(x_{0:2}|y_{1,2}) \propto p(x_{0:1}|y_1)\, p(x_2|x_1)\, p(y_2|x_2)$

weight: $\omega_2 = \frac{p(x_{0:2}|y_{1,2})}{q(x_{0:2}|y_{1,2})} \propto \underbrace{\frac{p(x_{0:1}|y_1)}{q(x_0)\, q(x_1|x_0, y_1)}}_{=\omega_1} \; \frac{p(x_2|x_1)\, p(y_2|x_2)}{q(x_2|x_1, y_2)}$

$$\omega_2 = \omega_1 \frac{p(x_2|x_1)\, p(y_2|x_2)}{q(x_2|x_1, y_2)}$$

Step ℓ:

Proposal : $q(x_{0:\ell}|y_{1:\ell}) \propto q(x_0) \prod_{j=1}^{\ell} q(x_j|x_{j-1}, y_j)$

target : $p(x_{0:\ell}|y_{1:\ell}) \propto p(x_{0:\ell-1}|y_{1:\ell-1})\, p(x_\ell|x_{\ell-1})\, p(y_\ell|x_\ell)$

weight: $$w_\ell = w_{\ell-1} \frac{p(x_\ell|x_{\ell-1})\, p(y_\ell|x_\ell)}{q(x_\ell|x_{\ell-1}, y_\ell)}$$

This is a "working" algorithm :

↳ we can evaluate target distribution (sequentially)

↳ we can build up the weights.

How to evaluate the target?

$$x_n = f(x_{n-1}) + v_n, \quad v_n \sim \mathcal{N}(0, Q)$$
$$y_n = h(x_n) + \eta_n, \quad \eta_n \sim \mathcal{N}(0, R)$$

$$p(x_n|x_{n-1}) \propto \exp\left(-\frac{1}{2}(x_n - f(x_{n-1}))^T Q^{-1}(x_n - f(x_{n-1}))\right)$$
$$p(y_n|x_n) \propto \exp\left(-\frac{1}{2}(y_n - h(x_{n-1}))^T R^{-1}(y_n - h(x_n))\right)$$

Algorithm: SIS → Sequential importance sampling ~~with resampling~~

At step k: $\{x^j_{0:k-1}, w_{k-1}\} \sim p(x_{0:k-1} | y_{1:k-1})$

ensemble of trajectories $\quad j = 1 \dots Ne.$

$\{x^j_{k-1}, w_{k-1}\} \sim p(x_{k-1} | y_{1:k-1})$

ensemble of states

Sample: $x^j_k \sim q(x | x^j_{k-1}, y_k)$

weight: $w_k \propto w_{k-1} \dfrac{p(x^j_k | x^j_{k-1})\, p(y_k | x^j_k)}{q(x^j_k | x^j_{k-1}, y_k)}$

new ensemble: $x^j_{0:k} = \{x^j_{0:k-1} ; x^j_k\}$

$\{x^j_{0:k}, w^j_k\}$ → ensemble of trajectories $\sim p(x_{0:k} | y_{1:k})$

$\{x^j_k, w^j_k\}$ → ensemble of states $\sim p(x_k | y_{1:k})$

Algorithm: SIR → Sequential importance sampling with resampling.

"More Common" →

ensemble: $\{x^j_{0:k-1}\} \sim p(x_{0:k-1} | y_{1:k-1})$

$\{x^j_{k-1}\} \sim p(x_{k-1} | y_{1:k-1})$

Sample: $x^j_k \sim q(x | x^j_{k-1}, y_k)$

weight: $w^j_k \propto w^j_{k-1} \dfrac{p(x^j_k | x^j_{k-1})\, p(y_k | x^j_k)}{q(x^j_k | x^j_{k-1}, y_k)}$

RESAMPLE: $x^j_k \leftarrow$ Resampled ensemble

new ensemble: $x^j_{0:k} = \{x^j_{0:k-1} ; x^j_k\}$

$\{x^j_{0:k}\} \sim p(x_{0:k} | y_{1:k}) \; ; \; \{x^j_k\} \sim p(x_k | y_{1:k})$