Quantifying Uncertainty

Sai Ravela

M. I. T

Last Updated: Spring 2013

Particle Filters

- Applied to Sequential filtering problems
- Can also be applied to smoothing problems
- Solution via Recursive Bayesian Estimation
- Approximate Solution
- Can work with non-Gaussian distributions/non-linear dynamics
- Applicable to many other problems e.g. Spatial Inference

Notation

x_t, X_k : Models states in continuous and discrete space-time respectively.

x_t^T : True system state

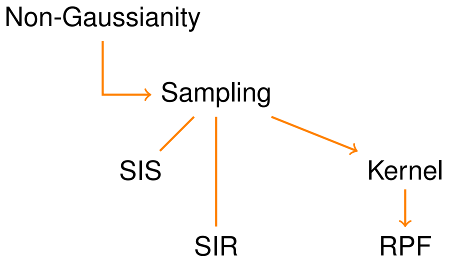
y_t, Y_k: Continous and Discrete measurements, respectively.

X_k^n : n^{th} sample of discrete vector at step k.

M: model, P: probability mass function.

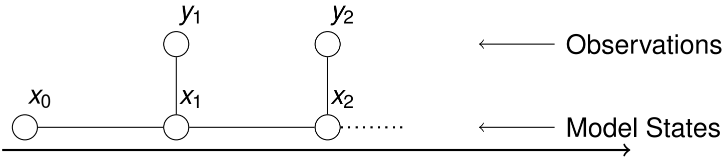
Q: Proposal Distribution, δ : kronecker or dirac delta function.

We follow Arulampalam et al.'s paper.



Sequential Filtering

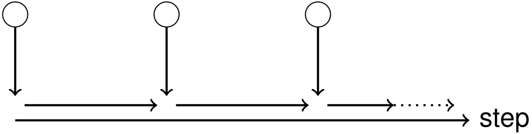
Recall: Ensemble Kalman filter & Smoother



We are interested in studying the evolution of $y_t \in f(x_t^T)$, observed system, using a model with state x_t.

This means (in discrete time, discretized space):

$$P(X_k|Y_{1:k})$$



Can be solved recursively

$$P(X_k|Y_{1:k}) = \frac{P(X_k, Y_{1:k})}{P(Y_{1:k})}$$

Sequential Filtering via Recursive Bayesian Estimation

$Y_{1:k}$ is a collection of variables $Y_1 \ldots Y_k$
So:

$$P(X_k|Y_{1:k}) = \frac{P(X_k, Y_{1:k})}{P(Y_{1:k})}$$

$$= \frac{P(Y_k|X_k)P(X_k|Y_k)\cancel{P(Y_{1:k-1})}}{P(Y_k|Y_{1:k-1})\cancel{P(Y_{1:k-1})}}$$

$$= \frac{P(Y_k|X_k)P(X_k|Y_{1:k-1})}{P(Y_k|Y_{1:k-1})}$$

Contd.

$$P(X_k|Y_{1:k}) = \frac{\underbrace{\underbrace{P(Y_k|X_k)}_{2} \sum_{X_{k-1}} P(X_k|X_{k-1})P(X_{k-1}|Y_{1:k-1})}_{1}}{\underbrace{\sum_{X_k} \sum_{X_{k-1}} P(Y_k|X_k)P(X_k|X_{k-1})P(X_{k-1}|Y_{k-1})}_{3}}$$

1. From the Chapman-Kolmogorov equation
2. The measurement model/observation equation
3. Normalization Constant

When can this recursive master equation be solved?

Let's say

$$
\begin{aligned}
X_k &= F_k X_{k-1} + V_k \\
Z_k &= H_k X_k + \eta_k \\
v_k &= N(\cdot, P_{k|k}) \\
\eta_k &= N(0, R)
\end{aligned}
$$

Linear Gaussian$\rightarrow$ Kalman Filter

For non linear problems

Extended Kalman Filter, via linearization

Ensemble Kalman filter

- No linearization
- Gaussian assumption
- Ensemble members are “particles” that moved around in state space
- They represent the moments of uncertainty

How may we relax the Gaussian assumption?

If $P(X_k|X_{k-1})$ and $P(Y_k|X_k)$ are non-gaussian;

How do we represent them, let alone perform these integrations in (2) & (3)?

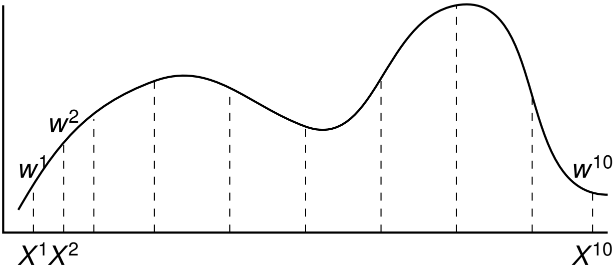
Particle Representation

Generically

$$P(X) = \sum_{i=1}^{N} w^i \delta(X - X^i)$$

pmf/pdf defined as a weighted sum
$\rightarrow$ Recall from Sampling lecture
$\rightarrow$ Response Surface Modeling lecture

Contd.



Even so,
Whilst $P(X)$ can be evaluated sampling from it may be difficult.

Importance Sampling

Suppose we wish to evaluate

$$\int_x f(x)P(x)dx \quad \text{(e.g. moment calculation)}$$

$$\int_x f(x)\frac{P(x)}{Q(x)}Q(x)dx, \qquad X^i \sim Q(x)$$

$$= \frac{1}{N}\sum_{i=1}^{N} f(x = X^i)w^i, \qquad w^i = \frac{P(x = X^i)}{Q(x = X^i)}$$

So:

Sample from $Q \equiv$Proposal distribution

Evaluate from $P \equiv$ the density

Apply importance weight $= w^i = \frac{P(X^i)}{Q(X^i)}$

Now let's consider

$$P(x) = \frac{\hat{P}(x)}{\int \widehat{P}(x)dx} = \frac{\widehat{P}(x)}{Z_p}$$

$$Q(x) = \frac{\hat{Q}(x)}{\int \widehat{Q}(x)dx} = \frac{\widehat{Q}(x)}{Z_q}$$

So:

$$\frac{1}{N}\frac{Z_q}{Z_p}\sum_{i=1}^{N} f(x = X^i)\widehat{w}^i$$

where

$$\widehat{w}^i = \frac{\widehat{P}(x = X^i)}{\widehat{Q}(x = X^i)}$$ These are un-normalized "mere potentials"

It turns out:

$$\frac{NZ_p}{Z_q} = \sum_i \hat{w}^i$$

$$\therefore \quad f(x)P(x)dx \cong \frac{\sum_{i=1}^{N} f(x = X^i)\hat{w}^i}{\sum_j^N \hat{w}^j}$$

$$\frac{\sum_i f(X^i)\hat{w}^i}{\sum_j \hat{w}^j} \quad \text{is just a "weighted sum"}$$

Where a proposal distribution was used to get around sampling difficulties and the importance weights manage all the normalization.

⇒It is important to select a good proposal distribution. Not one that focus on a small part of the state space and perhaps better than an uninformative prior.

Application of Importance Sampling to Bayesian Recursive Estimation

Particle Filter

$$P(X) \cong \frac{\sum_i \hat{w}^i \delta(X - X^i)}{\sum_j \hat{w}^j} = \sum_i w^i \delta(X - X^i)$$

w^i is normalized.

Let's consider again:

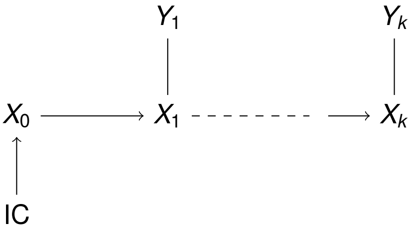
$$X_k = f(X_{k-1}) + V_k$$
$$Y_k = h(X_k) + \eta_k$$

A relationship between the observation and the state (measurement)

$\Rightarrow$ Additive noise, but can be generalized

Let's consider the joint distribution

$$P(X_{0:k}|Y_{1:K})$$



We may factor this distribution using particles

Chain Rule with Weights

$$P(X_{0:k}|Y_{1:k}) = \sum_{i=1}^{N} w^i \delta(X_{0:k} - X^i_{0:k})$$

$$w^i \equiv \frac{P(X^i_{0:k}|Y_{1:k})}{Q(X^i_{0:k}|Y_{1:k})}$$

And let's factor $P(X_{0:k}|Y_{1:k})$ as

$$\begin{aligned} P(X_{0:k}|Y_{1:k}) &= \frac{P(Y_k|X_{0:k}, Y_{1:k-1})P(X_{0:k}|Y_{1:k-1})}{P(Y_k|Y_{1:k-1})} \\ &= \frac{P(Y_k|X_k)P(X_k|X_{k-1})P(X_{k-1}|Y_{1:k-1})}{P(Y_k|Y_{1:k})} \end{aligned}$$

Proposal Distribution Properties

Suppose we pick

$$Q(X_{0:k}|Y_{1:k}) = Q(X_k|X_{0:k-1}, Y_{1:k})Q(X_{0:k-1}|Y_{1:k-1})$$

i.e. there is some kind of recursion on the proposal distribution. Further, if we approximate

$$Q(X_k|X_{0:k-1}, Y_{1:k}) = Q(X_k|X_{k-1}, Y_k)$$

i.e. there is a Markov property.

Recursive Weight Updates

Then we may have found an update equation for the weights.

$$\frac{P(X_{0:k}|Y_{1:k})}{Q(X_{0:k}|Y_{1:k})} = \frac{P(Y_k|X_k)P(X_k|X_{k-1})P(X_{0:k-1}, Y_{1:k-1})}{P(Y_k|Y_{1:k-1})Q(X_k|X_{k-1}, Y_k)Q(X_{0:k-1}|Y_{1:k-1})}$$

$$w_k^i = \frac{P(Y_k|X_k^i)P(X_k^i|X_{k-1}^i)}{Q(X_k^i|X_{k-1}^i, Y_k)P(Y_k|Y_{1:k-1})} \frac{P(X_{0:k-1}^i, Y_{1:k-1})}{Q(X_{0:k-1}^i, Y_{1:k-1})}$$

$$= \frac{P(Y_k|X_k^i)P(X_k^i|X_{k-1}^i)}{Q(X_k^i|X_{k-1}^i, Y_k)P(Y_k|Y_{1:k-1})} w_{k-1}^i$$

$$\propto \frac{P(Y_k|X_k^i)P(X_k^i|X_{k-1}^i)}{Q(X_k^i|X_{k-1}^i, Y_k)} w_{k-1}^i$$

The Particle Filter

In the filtering problem

$$P(X_k|Y_{1:k})$$

$$w_k^i \propto w_{k-1}^i \frac{P(Y_k|X_k^i)P(X_k^i|X_{k-1}^i)}{Q(X_k^i|X_{k-1}^i, Y_k)}$$

$$(So)\ P(X_k|Y_{1:k}) \cong \sum_{i=1}^{N} w_k^i \delta(X_k - X_k^i)$$

Where the $x_k^i \sim Q(X_k|X_{k-1}^i, Y_k)$
The method essentially draws particles from a proposal distribution and recursively update its weights.

$\Rightarrow$ No gaussian assumption

$\Rightarrow$ Neat

Algorithm Sequential Importance Sampling

Input: $\{X^i_{k-1}, w^i_{k-1}\}, \quad Y_k \quad i = 1 \dots N$

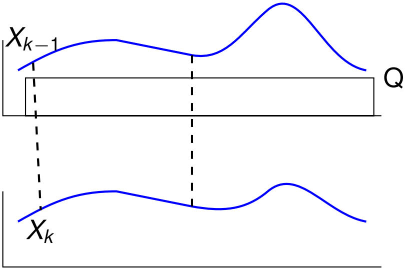
for: $i = 1 \dots N$

Draw: $X^i_k \sim Q(X_k | X^i_{k-1}, Y_k)$

$$w^i_k \propto w^i_{k-1} \frac{P(Y_k|X^i_k)P(X^i_k|X^i_{k-1})}{Q(X^i_k|X^i_{k-1}, Y_k)}$$

end

BUT The Problem



In a few intervals one particle will have a non negligible weight; all but one will have negligible weights!

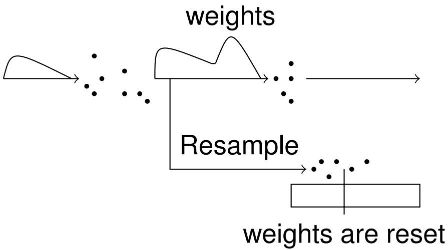
$$\widehat{N}_{eff} = \frac{1}{\sum_{i=1}^{N} (w_k^i)^2}$$

Contd.

$\widehat{N}_{eff} \equiv$ Effective Sample size
When $\widehat{N}_{eff} << N \rightarrow$ Degenaracy sets in

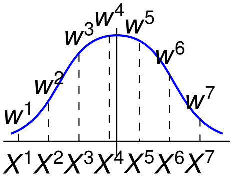
Resampling is a way to address this problem

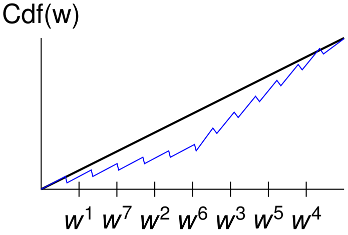
Main idea

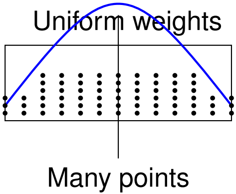


You can sample uniformly and set weights to obtain a representation. You can sample pdf to get particles and reset their weights.

Resampling algorithm







Resampling ⟶ Sample more probable states more

Algorithm

Input $\{X_k^i, w_k^i\}$

1. Construct cdf
   for $i = 2 : N \quad C_i \leftarrow C_{i-1} + w_k^i(sorted)$
2. Seed $u_1 \sim U[0, N^{-1}]$
3. for $j = 1 : N$
   $u_j = u_1 + \frac{1}{N}(j-1)$
   $i \leftarrow find(C_i \geq u_j)$
   $\hat{X}_k^j = X_k^i \quad w_k^j = \frac{1}{N}$
   Set Parent of $i^j \rightarrow i$

Contd.

So the resampling method can avoid degeneracy because it produces more samples for higher probability points

But Sample impoverishment may result; Too many samples too close $\rightarrow$ impoverishment or loss of diversity

$\Rightarrow$ MCMC may be a way out

Generic Particle filter

Input: $\{X_{k-1}^i, w_{k-1}^i\}, Y_k$

for $i = 1 : N$

$X_k^i \sim Q(X_k | X_{k-1}^i, Y_k)$

$w_k^i \leftarrow w_{k-1}^i \frac{P(Y_k|X_k^i)P(X_k^i|X_{k-1}^i)}{Q(X_k^i|X_{k-1}^i, Y_k)}$

end

$\eta = \sum_i w_k^i$

$w_k^i \leftarrow w_k^i / \eta$

$$\widehat{N}_{eff} = \frac{1}{\sum_{i=1}^N (w_k^i)^2}$$

If $\widehat{N}_{eff} < N_T$

$\{X_k, w_k^i\} \leftarrow$ Resample $\{X_k^i, w_k^i\}$

What is the optimal Q function?
we try to minimize $\sum_{i=1}^{N}(w_k^{*i})^2$
Then:

$$Q^*(X_k|X_{k-1}^i, Y_k) = P(X_k|X_{k-1}^i, Y_k)$$

$$= \frac{P(Y_k|X_k, X_{k-1}^i)P(X_k|X_{k-1}^i)}{P(Y_k|X_{k-1}^i)}$$

$$w_k^i \propto w_{k-1}^i \frac{P(Y_k|X_k^i)P(X_k^i|X_{k-1}^i)}{P(Y_k|X_k^i)P(X_k|X_{k-1}^i)} P(Y_k|X_{k-1}^i)$$

$$\propto w_{k-1}^i \underbrace{\int_{X_k} P(Y_k|X_k)P(X_k|X_{k-1}^i)dX_k}_{\text{Not easy to do!}}$$

Asymptotically:
$\widehat{Q} \sim P(X_k|X^i_{k-1}) \leftarrow$ Common choice $Q \equiv P(X_k|X^i_{k-1})$

Sometimes feasible to use proposal from process noise

Then

$w^i_K \propto w^i_{k-1} P(Y_k|X^i_k)$

If resampling is done at every step:

$w^i_k \propto p(Y_k|X^i_k)$

$(w^i_{k-1} \propto \frac{1}{N})$

SIR -Sampling Importance Resampling

Input $\{X^i_{k-1}, w^i_{k-1}\}, Y_k$

for $i = 1 : N$

$X^i_k \sim P(X_k | X^i_{k-1})$

$w^i_k = P(Y_k | X^i_k)$

end

$\eta = \sum_i w^i_k$

$w^i_k = w^i_k / \eta$

$\{x^i_k, w^i_k\} \leftarrow$ Resample $[\{X^i_k, w^i_k\}]$

Example

$$X_k = \frac{X_{k-1}}{2} + \frac{25X_{k-1}}{1+X_{k-1}^2} + 8\cos(1.2k) + v_{k-1}$$

$$Y_k = \frac{X_k^2}{w} + \eta_k$$

$\eta_k \sim N(0, R)$

$v_{k-1} \sim N(0, Q_{k-1})$

12.S990 Quantifying Uncertainty
Fall 2012