Importance Sampling and Particle Filtering

Namrata Vaswani, namrata@iastate.edu

I. Problem

A. State Space Model: HMM

Given a nonlinear state space model, satisfying the **Hidden Markov Model (HMM) assumptions**:

1) State sequence, $X_t, t = 0, 1, 2, ..$ is a Markov process, i.e. $p(x_t|x_{t-1}, past) = p(x_t|x_{t-1})$.
2) Observations, $Y_t, t = 1, 2, ..$ "conditionally independent given state at t", i.e. $p(y_t|x_t, past, future) = p(y_t|x_t)$ where $past \triangleq \{Y_{1:t-1}, X_{1:t-1}\}$, $future \triangleq \{Y_{t+1:T}, X_{t+1:T}\}$.
3) $p(x_0)$, $p(x_t|x_{t-1})$, $p(y_t|x_t)$ are known: often expressed using a system and observation model format.

B. Goal

1) Denote $x_{0:t} = \{x_0, x_1, ..x_t\}$ and $y_{1:t} = \{y_1, y_2, ..y_t\}$
2) Find a good approximation to $p(x_{0:t}|y_{1:t})$ or at least to $p(x_t|y_{1:t})$ referred to as the **posterior**. Goal is to estimate

$$I(f_t) = \int f_t(x_t)p(x_t|y_{1:t})dx_t \tag{1}$$

for any function of the state f_t

C. Some Other Techniques

1) If the state space model were **linear and Gaussian**, then $p(x_t|y_{1:t})$ is also Gaussian and its mean and variance are calculated using the Kalman filter.
2) **Nonlinear and Gaussian** : Extended Kalman Filter (EKF), Gaussian sum filter. Problem:
use first or second order Taylor series approximation to original nonlinear system, error in Jacobian estimates propagates over time, may be unstable in certain cases (error keeps increasing): loss of track. If one bad observation: loss of track, cannot come back
Cannot track heavily non-Gaussian or multimodal posteriors

3) **Multimodal systems:** Multiple Hypothesis tracker (MHT), Probabilistic Data Association Filters, Iterated Multiple Mode (IMM) etc

   Need to know or be able to estimate number of posterior modes
4) Unscented Kalman filter
5) Sigma point filters
6) Numerical integration techniques such as Quadrature filters - see Chapter 5 of [1]

   Practical only for small dimensional state space,
7) Grid-based filtering (discretizing the state space and then treating the system as a finite state HMM) - see [2]

   Practical only for small dimensional state space, small t
8) Markov Chain Monte Carlo

   Practical only for small dimensional state space, small t

II. Importance Sampling

A Particle Filter is a Sequential Monte Carlo method. It is a modification of the Sequential Importance Sampling method. Need to first understand (read first few pages of [3] or Chapter 1 of [1])

1) Monte Carlo sampling from a pdf $a(x)$
2) Importance sampling when a is known in closed form, but cannot sample from it. So sample from a q (satisfying certain assumptions given in [3] and also in [1]).
3) Importance sampling when $a(x) = \frac{\tilde{a}(x)}{\int \tilde{a}(z)dz}$, only $\tilde{a}$ is known in closed form, a is not (because the denominator cannot be evaluated analytically).

The third case is treated as follows:

$$\begin{aligned} I(f) = \int f(x)a(x)dx &= \frac{\int f(x)\tilde{a}(x)dx}{\int 1\tilde{a}(z)dz} \\ &= \frac{\int f(x)\frac{\tilde{a}(x)}{q(x)}q(x)dx}{\int 1\frac{\tilde{a}(z)}{q(z)}q(z)dz} \\ &\approx \sum_{i=1}^{N} f(x^i)\frac{\frac{\tilde{a}(x^i)}{q(x^i)}}{\sum_{j=1}^{N}\frac{\tilde{a}(x^i)}{q(x^i)}}, \quad \text{with} \quad x^i \sim q(x) \end{aligned} \tag{2}$$

or in other words (some misuse of notation),

$$a(x) \approx \sum_{i=1}^{N} \frac{\tilde{w}^i}{\sum_{j=1}^{N}\tilde{w}^j}\delta(x - x^i), \quad \tilde{w}^i \triangleq \frac{\tilde{a}(x^i)}{q(x^i)}, \quad \text{with} \quad x^i \sim q(x) \tag{3}$$

where δ denotes the Dirac delta function (that you learn in Signals and Systems). Note a more commonly used (and perhaps mathematically correct) notation is to write $a(dx) \approx \sum_{i=1}^{N} \frac{\tilde{w}^i}{\sum_{j=1}^{N} \tilde{w}^j} \delta_{x^i}(dx)$. The above is a biased estimate as explained in [3], [1]. But it is asymptotically (for $N \to \infty$) unbiased.

III. Sequential Importance Sampling (SIS)

Sequential Importance Sampling (SIS) to approximate $p(x_{0:t}|y_{0:t})$: uses the above idea with $x \triangleq x_{0:t}$ and $\tilde{a}(x_{0:t}) \triangleq p(x_{0:t}, y_{1:t})$ and $q(x_{0:t}) \triangleq \pi(x_{0:t}|y_{1:t})$. To get a recursive algorithm that can be applied at any t using only observations until t, we need to choose $q = \pi$ that satisfies

$$q(.) = \pi(x_{0:t}|y_{1:t}) = \pi(x_{0:t-1}|y_{1:t-1})\pi(x_t|x_{0:t-1}, y_{0:t}) \tag{4}$$

Using the HMM assumptions given above, $\tilde{a}$ can be re-written as

$$\tilde{a}(.) = p(x_{0:t}, y_{1:t}) = p(y_{1:t-1}, x_{0:t-1})p(y_t|x_t)p(x_t|x_{t-1}) \tag{5}$$

Thus

$$\tilde{w}_t^i = \frac{\tilde{a}(.)}{q(.)} = \frac{p(x_{0:t}^i, y_{1:t})}{\pi(x_{0:t}^i|y_{1:t})} = \tilde{w}_{t-1}^i \frac{p(y_t|x_t^i)p(x_t^i|x_{t-1}^i)}{\pi(x_t^i|x_{0:t-1}^i, y_{0:t})}, \quad \text{with} \quad x_t^i \sim \pi(x_t^i|x_{0:t-1}^i, y_{0:t}) \tag{6}$$

This gives the SIS algorithm. See the algorithm on Page 6 of [3].

A. *Different choices for π:*

1) The Bayesian Boostrap filter [4] used $\pi(x_t|x_{0:t-1}, y_{0:t}) = p(x_t|x_{t-1})$ and $\pi(x_0) = p(x_0)$. This is also the most commonly used implementation. Then the recursion for weights is

$$\tilde{w}_t^i = \tilde{w}_{t-1}^i p(y_t|x_t^i) \tag{7}$$

2) The optimal importance density (minimizes variance of weights given previous states and all observations till t) is to use $\pi(x_t|x_{0:t-1}, y_{0:t}) = p(x_t|x_{t-1}, y_t)$ and $\pi(x_0) = p(x_0)$. Actually note that $p(x_t|x_{t-1}, y_t) = p(x_t|x_{0:t-1}, y_{0:t})$ because of HMM assumptions. Often cannot compute this, but if we know it is unimodal, then can try to use Gaussian approximations to it (details in [3]). In this case

$$\tilde{w}_t^i = \tilde{w}_{t-1}^i \frac{p(y_t|x_t^i)p(x_t^i|x_{t-1}^i)}{p(x_t^i|x_{t-1}^i, y_t)} \tag{8}$$

If using exact $p(x_t|x_{t-1}, y_t)$, then $\tilde{w}_t^i = \tilde{w}_{t-1}^i p(y_t|x_{t-1}^i)$.

3) Many other choices exist in literature.

IV. Particle Filter

SIS as described above becomes impractical after a few time steps, since the variance of $\tilde{w}_t^i$ become very large (order of magnitude difference between largest and smallest weights), consequently in the normalized weights only one weight is significantly non-zero, others are practically zero. This is called “particle inefficiency” - the particles with practically zero weights are being “wasted” - not approximating the posterior in high probability regions, while there are very few (often just one) particles in the high probability region.

Particle filter (or Bayesian Boostrap filter) is Sequential Importance Sampling along with a Resampling step which is described in detail in Section 3 of [3], and it addresses the above issue. See the algorithm on page 13 of [3].

Under mild assumptions, the particle filtering estimate of the posterior distribution of x_t given $y_{1:t}$ can be shown to converge weakly to the true posterior as the number of particles, $N \to \infty$ (see [1], chapters 2 and 3, if interested).

V. Variants of the basic Particle Filter (PF)

1) Rao Blackwellized PF: runs a Kalman Filter for a linear/Gaussian subsystem. See [2].
2) Regularized PF
3) Unscented PF, Sigma point PF
4) PF with approximating the optimal Importance Sampling distribution in different ways. See [3], [2].
5) Auxiliary PF

References

[1] A. Doucet, N. deFreitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*, Springer, 2001.

[2] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, “A tutorial on particle filters for on-line non-linear/non-gaussian bayesian tracking,” *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

[3] A. Doucet, “On sequential simulation-based methods for bayesian filtering,” Tech. Rep. CUED/F-INFENG/TR. 310, Cambridge University Department of Engineering, 1998.

[4] N.J. Gordon, D.J. Salmond, and A.F.M. Smith, “Novel approach to nonlinear/nongaussian bayesian state estimation,” *IEE Proceedings-F (Radar and Signal Processing)*, pp. 140(2):107–113, 1993.