Particle Filtering

Dynamical systems are often modeled using a state-space framework, where the phenomenon of interest is viewed as the unknown state of the system, and noisy measurements of the state are available to an observer. The state-space model is often assumed to be linear:

$$x_{t+1} = F_t x_t + G_t u_t \tag{1}$$
$$y_t = H_t x_t + v_t \tag{2}$$

where $t = 0, 1, 2, \cdots$ is the time index, $x_t \in \mathbb{R}^m$ is the *state vector*, $y_t \in \mathbb{R}^k$ is the *measurement vector*, and $u_t \in \mathbb{R}^p$ and $v_t \in \mathbb{R}^q$ are independent noise vectors. In these equations, F_t, G_t, and H_t are known matrices. The second-order statistics of the noise vectors U_t, V_t and the initial state X_0 are known. The well-known Kalman filter recursively produces the so-called linear MMSE estimator [1] of X_t and the linear MMSE predictor of X_{t+1} given all the data $Y_{0:t} = (Y_0, Y_1, \cdots, Y_t)$ available at time t. If the noise sequences and the initial state are Gaussian, the estimates are also MMSE estimates.

In many problems, the noise in non-Gaussian and/or the linear model (1) (2) is invalid. The performance of the Kalman filter can be disastrous in such cases – e.g., the filter may fail to track its target. The *extended Kalman filter* can cope with some nonlinear models by using linearization of the state equation around the predicted state [1]. However this filter may still perform poorly if the linearization is crude and/or the noise statistics are strongly non-Gaussian.

This has motivated the development of algorithms for nonlinear state-space models in which optimal MMSE estimators and predictors are produced by a *nonlinear recursive filter* [2]. The state-space model is as follows:

$$x_{t+1} = f_t(x_t, u_t) \tag{3}$$
$$y_t = h_t(x_t, v_t) \tag{4}$$

where the mappings $f_t : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}^m$ and $h_t : \mathbb{R}^m \times \mathbb{R}^q \to \mathbb{R}^k$ are assumed to be known, as are the pdf's $p_{U_t}(u_t)$ and $p_{V_t}(v_t)$ for the noise vectors, and the pdf $p_{X_0}(x_0)$ of the initial state. The noise sequences are also temporally independent.

Estimation problem: estimate X_t given $Y_{0:t}$. We denote this estimator by $\hat{X}_{t|t}$.
Prediction problem: predict X_{t+1} given $Y_{0:t}$. We denote this predictor by $\hat{X}_{t+1|t}$.

In Sec. 1 we derive a recursive Bayesian filter for this nonlinear dynamical system. This solution requires numerical evaluation of integrals, and this is done using a sequential importance sampling method (the particle filter, *aka* the bootstrap filter) [2], as shown in Sec. 2.

Nowadays particle filtering methods are widely applied to problems arising in communications, image and video processing, and statistics [3, 4, 5, 6].

[1] That is, the best *linear filter* in the MMSE sense.

1 Bayesian Recursive Filtering

The Bayesian approach requires a cost function $C(x_t, \hat{x}_t)$ and the posterior pdf's $p(x_t|y_{0:t})$ and $p(x_{t+1}|y_{0:t})$. For MMSE estimation, we have $C(x_t, \hat{x}_t) = \|x_t - \hat{x}_t\|^2$. The optimal estimator and predictor are given by the conditional means

$$\hat{X}_{t|t} = \int_{\mathbb{R}^m} x_t \, p(x_t|Y_{0:t}) \, dx_t, \tag{5}$$

$$\hat{X}_{t+1|t} = \int_{\mathbb{R}^m} x_{t+1} \, p(x_{t+1}|Y_{0:t}) \, dx_{t+1}. \tag{6}$$

The posterior pdf's can *in principle* be evaluated recursively using the following two-step procedure.

Step 1: Prediction. We can express $p(x_{t+1}|y_{0:t})$ in terms of $p(x_t|y_{0:t})$:

$$\begin{aligned}
p(x_{t+1}|y_{0:t}) &= \int_{\mathbb{R}^m} p(x_{t+1}|x_t, y_{0:t}) \, p(x_t|y_{0:t}) \, dx_t \\
&= \int_{\mathbb{R}^m} p(x_{t+1}|x_t) \, p(x_t|y_{0:t}) \, dx_t \\
&= \int_{\mathbb{R}^m} \left[\int_{\mathbb{R}^p} 1\{x_{t+1} = f_t(x_t, u_t)\} \, p(u_t) \, du_t \right] p(x_t|y_{0:t}) \, dx_t
\end{aligned}$$

where in the second line we have used the fact that $Y_{0:t} \to X_t \to X_{t+1}$ forms a Markov chain.

Step 2: Update. Next we express $p(x_t|y_{0:t})$ in terms of $p(x_t|y_{0:t-1})$:

$$\begin{aligned}
p(x_t|y_{0:t}) &= p(x_t|y_t, y_{0:t-1}) \\
&= \frac{p(y_t|x_t, y_{0:t-1}) \, p(x_t|y_{0:t-1})}{p(y_t|y_{0:t-1})} \\
&= \frac{p(y_t|x_t, y_{0:t-1}) \, p(x_t|y_{0:t-1})}{\int_{\mathbb{R}^m} p(y_t|x_t, y_{0:t-1}) \, p(x_t|y_{0:t-1}) \, dx_t} \\
&= \frac{p(y_t|x_t) \, p(x_t|y_{0:t-1})}{\int_{\mathbb{R}^m} p(y_t|x_t) \, p(x_t|y_{0:t-1}) \, dx_t} \\
&= \frac{[\int_{\mathbb{R}^q} 1\{y_t = h_t(x_t, v_t)p(v_t) \, dv_t] \, p(x_t|y_{0:t-1})}{\int_{\mathbb{R}^m} [\int_{\mathbb{R}^q} 1\{y_t = h_t(x_t, v_t)p(v_t) \, dv_t] \, p(x_t|y_{0:t-1}) \, dx_t}
\end{aligned}$$

where in the second line follows from Bayes' rule, and the fourth line from the fact that $Y_{0:t-1} \to X_t \to Y_t$ forms a Markov chain.

As discussed in the previous lecture, due to the curse of dimensionality, evaluation of such integrals (one for each x_t!) using deterministic quadrature methods is generally computationally infeasible. The modern approach to recursive MMSE filtering is therefore to use stochastic integration methods.

2 Particle Filter

The particle filter, also called bootstrap filter, is a sequential Monte Carlo method which outputs a stochastic approximation to $\hat{X}_{t|t}$ and $\hat{X}_{t+1|t}$ in (5) and (6).

The first idea is to approximate the conditional-mean estimator and predictor of (5) and (6) with sample averages:

$$\tilde{X}_{t|t} = \frac{1}{N}\sum_{i=1}^{N} X_t(i) \tag{7}$$

$$\tilde{X}_{t+1|t} = \frac{1}{N}\sum_{i=1}^{N} X^*_{t+1}(i) \tag{8}$$

where the N samples $X_t(i)$, $1 \leq i \leq N$, are drawn iid from the posterior pdf $p(x_t|Y_{0:t})$, and similarly, the N samples $X^*_{t+1}(i)$, $1 \leq i \leq N$, are drawn iid from the posterior pdf $p(x_{t+1}|Y_{0:t})$.

The second idea (to avoid computation of the integrals giving the posterior pdf's from which the samples are drawn) is to reuse the samples $X_t(i)$, $1 \leq i \leq N$, to generate $X^*_{t+1}(i)$, $1 \leq i \leq N$ and $X_{t+1}(i)$, $1 \leq i \leq N$. Hence the algorithm propagates a set of particles over time. This is done as follows. Assume we have N iid samples $X_t(i)$, $1 \leq i \leq N$, drawn from $p(x_t|Y_{0:t})$ and N iid noise samples $U_t(i)$, $1 \leq i \leq N$ drawn from $p(u_t)$.

Step 1: Prediction. Define

$$X^*_{t+1}(i) = f_t(X_t(i), U_t(i)), \quad 1 \leq i \leq N.$$

By construction, these samples are iid with marginal distribution $p(x_{t+1}|Y_{0:t})$.

Step 2: Update. Upon receiving a new measurement y_{t+1}, evaluate the *importance weights*

$$q_i = \frac{p(y_{t+1}|X^*_{t+1}(i))}{\sum_{j=1}^{N} p(y_{t+1}|X^*_{t+1}(j))}, \quad 1 \leq i \leq N.$$

Resample N times from the set $\{X_t(i)\}$ with respective probabilities $\{q_i\}$, obtaining updated samples $\{X_{t+1}(j)\}$ with probabilities

$$Pr[X_{t+1}(j) = X_t(i)] = q_i, \quad 1 \leq i, j \leq N.$$

By the weighted bootstrap theorem (see appendix), the distribution of the resampled $\{X_{t+1}(j)\}$ converges to the desired posterior

$$p(x_{t+1}|y_{0:t+1}) \propto p(y_{t+1}|x_{t+1})\, p(x_{t+1}|y_{0:t})$$

where the posterior $p(x_{t+1}|y_{0:t+1})$, the "likelihood function" $p(y_{t+1}|x_{t+1})$ and the "prior" $p(x_{t+1}|y_{0:t})$ play the role of $h(\cdot)$, $l(\cdot)$ and $g(\cdot)$ in (9).

Choice of N. The number of particles should be "large enough" so that the stochastic approximations to conditional expectations are accurate, but how large should N be? The biggest problem is the potential weak overlap of the "likelihood" $p(y_{t+1}|x_{t+1})$ and the "prior" $p(x_{t+1}|y_{0:t})$, which can cause the same samples to be reused many times, leading to an *effective* number of samples that is much smaller than N. This problem becomes more significant in high dimensions. Methods have been proposed in the statistics literature that alleviate this problem to some extent.

A Weighted Bootstrap

Given an arbitrary positive function $l(x)$ over $\mathbb{R}^m$ and a set of samples $\{X(i),\, 1 \le i \le N\}$ drawn iid from a pdf $g(x)$, we would like to produce iid samples $\{z(i),\, 1 \le i \le N\}$ from the pdf

$$h(x) = \frac{g(x)l(x)}{\int g(x)l(x)\,dx}. \tag{9}$$

Define the *importance weights*

$$q_i = \frac{l(X(i))}{\sum_{j=1}^{N} l(X(j))}, \quad 1 \le i \le N$$

and the empirical pdf

$$\tilde{h}_N(x) = \sum_{i=1}^{N} q_i \delta(x - X(i))$$

where $\delta(\cdot)$ is the Dirac impulse.

The weighted bootstrap resamples the set $\{X(i),\, 1 \le i \le N\}$ with respective probabilities $\{q_i\}$, i.e., for each resample Z we have

$$Pr[Z = X(i)] = q_i, \quad 1 \le i \le N.$$

Proposition: Z converges in distribution to h.

Proof. For notational simplicity we give the proof only in the 1-D case ($m = 1$). We have

$$\begin{aligned} Pr[Z \le z] &= \sum_{i=1}^{N} q_i\, 1_{\{X(i) \le z\}} \\ &= \frac{\frac{1}{N}\sum_{i=1}^{N} l(X(i))\, 1_{\{X(i) \le z\}}}{\frac{1}{N}\sum_{i=1}^{N} l(X(i))} = \frac{N(z)}{N(\infty)}. \end{aligned}$$

Since X_i are drawn iid from $g(x)$, it follows from the weak law of large numbers that $N(z)$ above converges in probability to

$$\mathbb{E}_g[N(z)] = \mathbb{E}_g[l(X)\, 1_{\{X \le z\}}] = \int_{-\infty}^{z} g(x)l(x)\,dx.$$

Likewise, the denominator $N(\infty)$ converges in probability to $\int_{-\infty}^{\infty} g(x)l(x)\,dx$. Hence

$$\lim_{N\to\infty} Pr[Z \leq z] = \frac{\int_{-\infty}^{z} g(x)l(x)\,dx}{\int_{-\infty}^{\infty} g(x)l(x)\,dx} = \int_{-\infty}^{z} h(x)\,dx$$

which proves the claim. □

References

[1] B. D. Anderson and J. B. Moore, *Optimal Filtering*, Prentice-Hall, NJ, 1979.

[2] N. Gordon, D. Samond, and A. F. M. Smith, "Novel Approach to Nonlinear and Non-Gaussian Bayesian State Estimation," *IEE Proceedings-F*, Vol. 140, pp. 107—113, Apr. 1993.

[3] J. Liu and R. Chen, "Sequential Monte Carlo methods for Dynamical Systems," *J. Amer. Stat. Assoc.*, Vol. 93, pp. 1032—1044, Sep. 1998.

[4] A. Doucet, N. de Freitas and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*, Springer, NY, 2001.

[5] M. S. Arulampalam, S. Maskell N. Gordon, and T. Clapp, "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking," *IEEE Trans. Signal Processing*, Vol. 50, No. 2, pp. 174—188, Feb. 2002.

[6] P. M. Djuric, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Miguez, "Applications of particle filtering to selected problems in communications: A review and new developments," *Signal Processing Magazine*, Vol. 20, No. 5, pp. 19—38, 2003.

[7] A. F. M. Smith and Gelfand, "Bayesian Statistics Without Tears: A Sampling-Resampling Perspective," *The American Statistician*, Vol. 46, pp. 80—88, May 1992.