ECE 830 Fall 2011 Statistical Signal Processing

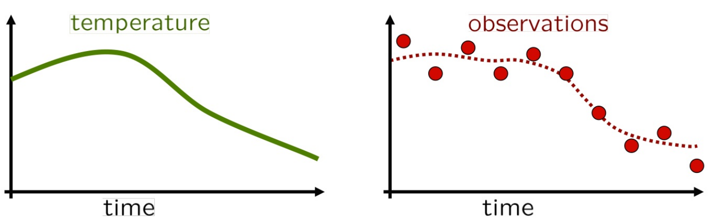
instructor: R. Nowak

Lecture 22: Dynamic Filtering

1 Dynamic Filtering

In many applications we want to track a time-varying (dynamic) phenomenon.

Example 1 *Tracking temperature or humidity in a museum room with an inaccurate device.*



Key: Temperature changes slowly with time so we should be able to average across time to obtain better estimates. How to do this? Model dynamics of temperature changes and noise/uncertainties in measurement.

2 Dynamical State Equation (Prior)

Let x_1, x_2, ... denotes quantity ("state") of interest. The state is changing over time and we will model this variation stochastically as follows. The state at time n depends causally on the past. Let $p(x_n|x_{n-1}, x_{n-2}, \ldots, x_1)$ denote the conditional distribution of the state at time n given all the past states. This distribution is a n-variate function, and as n grows it becomes more and more complex (to specifiy, to compute, etc). A reasonable simplifying assumption is to assume that the probability distribution of the state at time n depends only on value of the state at time $n-1$, a so-called *Markovian* assumption,

$$p(x_n|x_{n-1}, \ldots, x_1) = p(x_n|x_{n-1}) \ .$$

Note that $p(x_n|x_{n-1})$ is bivariate and therefore much simpler than the general causal model. To define the state process we must to specify

(a) $p(x_1)$, the "initial state" distribution

(b) $p(x_n|x_{n-1})$, $n = 2, 3, \ldots$, the state transition probability density functions

This is illustrated in the following example.

Example 2 *Santa Tracker On December 25th legend has it that Santa Claus makes his way around the globe, delivering toys to all the good girls and boys. Tracking Santa's delivery trip has attracted considerable* *interest by the signal processing research community in recent years, see* `http://www.noradsanta.org/`*. Here is a simple approach to the problem.*

$$\begin{aligned} x(t) &= \textit{Santa's position at time } t \textit{ on Christmas Eve} \\ \frac{\partial x(t)}{\partial t} &= v(t), \textit{ velocity} \end{aligned}$$

We can sample Santa's position once every second, producing a sequence of position values $x_1, x_2, \ldots$. *His velocity is also represented by a discrete-time process* $v_1, v_2, \ldots$ *We use the following model for Santa's dynamics:*

$$\begin{bmatrix} x_{n+1} \\ v_{n+1} \end{bmatrix} = \begin{bmatrix} 1 & \Delta \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_n \\ v_n \end{bmatrix} + \begin{bmatrix} 0 \\ \sigma^2 \end{bmatrix} u_n, \; u_n \sim \mathcal{N}(0,1), \Delta \textit{ small}$$

Also, Santa's initial position is the North Pole, denoted by x_0. *So we take* $p(x_1) = \delta(x_1 - x_0)$. *In words, Santa's position at time* x_{n+1} *is equal to his position at time* n *plus a small step proportional to his velocity. His velocity is modeled as a Gaussian white noise process, representing the fact that he randomly speeds up and slows down as he makes his stops around the world.*

3 Observation Model (Likelihood)

Usually we cannot observe x_n directly. Instead we observe $z_1, z_2, \ldots,$ which are noisy and/or indirect measurements related to the states.

Example 3 *Here are a few examples of observation processes.*

$$\begin{aligned} z_n &= x_n + w_n \;,\; w_n \sim \mathcal{N}(0, \sigma^2) \;, \textit{ simple signal+noise model} \\ z_n &= Ax_n + w_n \;, \textit{ where } A \textit{ is a matrix representing, for example, a blur} \\ z_n &= f(x_n) + w_n \;, f \textit{ is a non-linear function} \end{aligned}$$

Let $p(z_n|x_n)$ denote the likelihood of x_n based on observation z_n. We can combine the likelihoods and the priors $p(x_n|x_{n-1})$ to compute the posterior distribution of $x = (x_1, \ldots, x_n)$ given $z = (z_1, \ldots, z_n)$

$$p(x|z) \;\propto\; p(z|x)p(x) \;=\; \prod_{i=1}^{n} p(z_i|x_i)p(x_i|x_{i-1}) \;.$$

The posterior can be computed efficiently in an incremental fasion by exploiting Markovian structure of state transitions (prior). This incremental procedure is called *Density Propagation.*

4 Density Propagation

Density Propagation is an incremental procedure for efficiently computing $p(x_n|z_1, \ldots, z_n)$. First let's establish some notation.

<u>Prior:</u>

$$S_n(x_n|x_{n-1}) := p(x_n|x_{n-1}), \quad P_1(x_1) = p(x_1)$$

<u>Likelihood:</u>

$$L_n(z_n|x_n) := p(z_n|x_n)$$

<u>Posterior:</u>

$$F_n(x_n) := p(x_n|z_1, \ldots, z_n)$$

Prediction:

$$P_n(x_n) := p(x_n|z_1, \ldots, z_{n-1})$$

$P_n(x_n)$ is the prediction of the value of x_n using only observations up to time $n-1$, and this will play a key role in the Density Propagation algorithm.

4.1 Density Propogation Algorithm

n = 1:
predict x_1:

$$x_1 \sim p_1(x_1)$$

observe z_1 and
compute posterior:

$$F_1(x_1) = p(x_1|z_1) = \frac{p(z_1|x_1)p(x_1)}{p(z_1)} \propto L_1(z_1|x_1)p_1(x_1)$$

n = 2:
predict x_2:

$$\begin{aligned}
p(x_1, x_2|z_1) &= \frac{p(x_1, x_2, z_1)}{p(z_1)} \\
&= \frac{p(x_2|x_1, z_1)p(x_1|z_1)p(z_1)}{p(z_1)} \\
&= p(x_2|x_1)F_1(x_1) \\
&= S_2(x_2|x_1)F_1(x_1) \\
p(x_2|z_1) &= \int S_2(x_2|x_1)F_1(x_1)\mathrm{d}x \\
&=: P_2(x_2)
\end{aligned}$$

observe z_2 and
compute posterior:

$$\begin{aligned}
F_2(x_2) &= p(x_2|z_1, z_2) \\
&= \frac{p(x_2, z_1, z_2)}{p(z_1, z_2)} \\
&= \frac{p(z_2|x_2)p(x_2|z_1)p(z_1)}{p(z_1, z_2)} \\
&\propto L_2(z_2|x_2)P_2(x_2)
\end{aligned}$$

at time step n:
predict x_n:

$$\begin{aligned}
P_n(x_n) &= p(x_n|z_1, \ldots, z_{n-1}) \\
&= \int S_n(x_n|x_{n-1})F_{n-1}(x_{n-1})\mathrm{d}x_{n-1}
\end{aligned}$$

observe z_n and
compute posterior:

$$\begin{aligned}
F_n(x_n) &= p(x_n|z_1, \ldots, z_n) \\
&\propto L_n(z_n|x_n)P_n(x_n)
\end{aligned}$$

4.2 Block Diagram

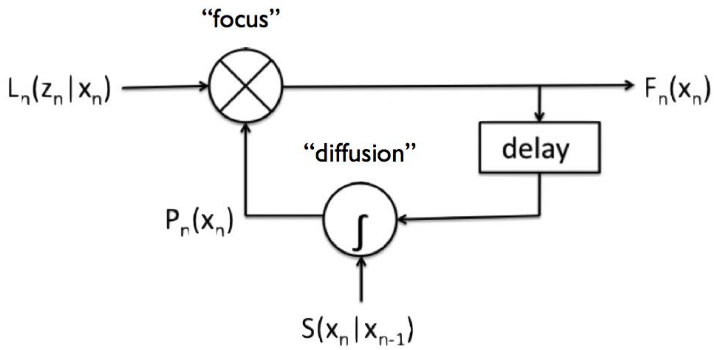


Figure 1: Block diagram of dynamic filtering.

4.2.1 Filtering

$$F_n(x_n) = L_n(z_n|x_n)P_n(x_n)$$

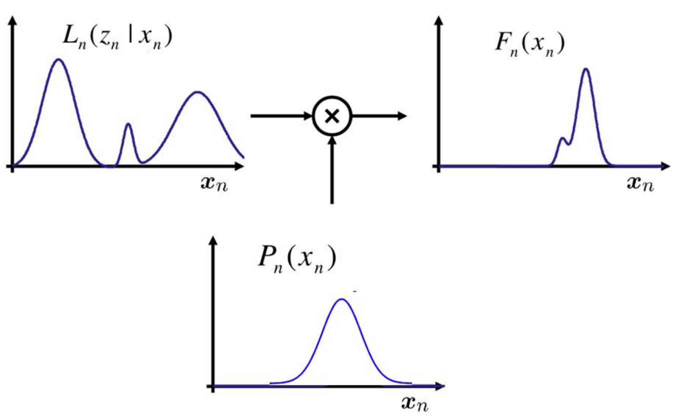


Figure 2: The filtering or "focus" portion of the dynamical filtering block diagram.

4.2.2 Prediction

$$P_{n+1}(x_{n+1}) = \int S_n(x_{n+1}|x_n)F_n(x_n)\mathrm{d}x_n$$

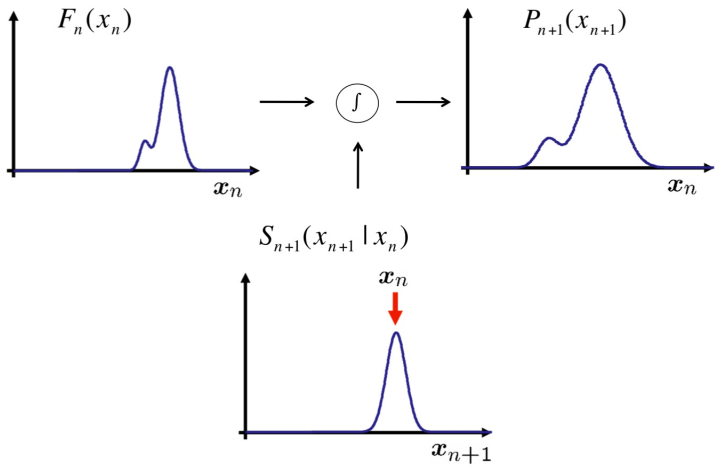


Figure 3: The prediction or “diffusion” portion of the dynamical filtering block diagram.

5 Estimating x_n

We have many possibilities. Given,

$$F_n(x_n) = p(x_n|z_1, \dots, z_n)$$

We can minimize various risk functions based on a loss and the posterior distribution F_n.

ℓ_2:

$$\begin{aligned} \widehat{x}_n &= \arg\min_{\widetilde{x}} \mathbb{E}_{F_n}[(x_n - \widetilde{x})^2] \\ &= \int x_n F_n(x_n) \mathrm{d}x_n \end{aligned}$$

ℓ_1:

$$\widehat{x}_n = \arg\min_{\widetilde{x}} \mathbb{E}_{F_n}[|x_n - \widetilde{x}|]$$

$\ell_{0/1}$:

$$\widehat{x}_n = \arg\max_{x} F_n(x_n)$$