A Dummies Guide to the Kalman filter and Sequential Monte Carlo methods (aka Particle filters)

Suhasini Subba Rao

July, 2004

A simple state space model

We observe $\{y_t : t = 1, \ldots, T\}$, where y_t satisfies

$$y_{t+1} = x_{t+1} + e_{t+1}, \quad \text{observation equation}$$

and $\{x_t\}$ is not observed and satisfies

$$x_{t+1} = x_t + w_{t+1} \quad \text{system equation.}$$

Motivation We can think $x_{t+1} = x(t + \delta t)$
x(t) a smooth function in time:

$$\begin{aligned} x(t+\delta t) &= x(t) + \text{ higher order terms} \\ \Rightarrow x_{t+1} &= x_t + w_{t+1}. \end{aligned}$$

(West and Harrison, 1997)

The general nonlinear state-space model

We observe $\{y_t : t = 1, \ldots, T\}$, where y_t satisfies

$$\begin{aligned} y_t &= h_t(x_t, e_t) \quad \text{observation equation} \\ x_t &= f_t(x_{t-1}, w_t) \quad \text{system equation} - \quad x_t \text{ not observed.} \end{aligned}$$

The distribution functions $p(x_t|x_{t-1})$ and $p(y_t|x_t)$ are known.
$\underline{Y}_t = (y_1, \ldots, y_t)$.

Object: Recursively evaluate $p(x_t|\underline{Y}_{t-1})$ and $p(x_t|\underline{Y}_t)$

- $p(x_t|\underline{Y}_{t-1})$ is the predictor distribution.
- $p(x_t|\underline{Y}_t)$ is the filter distribution.

The generalised Kalman equations

The predictor:

$$p(x_t|\underline{Y}_{t-1}) = \int p(x_{t-1}|\underline{Y}_{t-1}) p(x_t|x_{t-1}) dx_{t-1}$$

The filter

$$p(x_t|\underline{Y}_t) = \frac{p(y_t|x_t) p(x_t|\underline{Y}_{t-1})}{p(y_t|\underline{Y}_{t-1})}$$

We can ping-pong between the filter and predictor and finally evaluate the smoother: **The smoother**

$$p(x_t|\underline{Y}_T) = p(x_t|\underline{Y}_t) \int \frac{p(x_{t+1}|\underline{Y}_T) p(x_{t+1}|x_t)}{p(x_{t+1}|\underline{Y}_t)} dx_{t+1}$$

Despite simplicity of equations, in practice difficult to evaluate.

Example: The Kalman filter

The Gaussian linear (classical) model:

$$
\begin{aligned}
x_t &= F_t x_{t-1} + G_t w_t \quad w_t \sim N(0, Q) \\
y_t &= H_t x_t + e_t \quad e_t \sim N(0, R_t)
\end{aligned}
$$

F_t, G_t, H_t, Q, R_t assumed known. Assume we have $x_{t-1|t-1}$ and $V_{t-1|t-1}$:

- *Predictor equation*:

$$
\begin{aligned}
\mathbf{E}(x_t|\underline{Y}_{t-1}) &= x_{t|t-1} = F_t x_{t-1|t-1} \\
\mathrm{var}(x_t|\underline{Y}_{t-1}) &= V_{t|t-1} = F_t V_{t-1|t-1} F_t^T + G_t Q G_t^T,
\end{aligned}
$$

and the predictor distribution is

$$p(x_t|\underline{Y}_{t-1}) \sim N(x_{t|t-1}, V_{t|t-1}).$$

- *Filter*:

$$
\begin{aligned}
K_t &= V_{t|t-1} H_t^T \left(H_t V_{t|t-1} H_t^T + R_t\right)^{-1} \quad \text{(the Kalman Factor)} \\
\mathbf{E}(x_t|\underline{Y}_t) &= x_{t|t} = x_{t|t-1} + K_t(y_t - H_t x_{t|t-1}) \\
\mathrm{var}(x_t|\underline{Y}_t) &= V_{t|t} = (I - K_t H_t) V_{t|t-1}
\end{aligned}
$$

then the filter distribution is

$$p(x_t|\underline{Y}_t) \sim N(x_{t|t}, V_{t|t}).$$

- The smoother filter can be similarly defined (omitted here).

Proof, for example, can be found in (Kitagawa, 1996, p85)

Problem: For the Gaussian linear state-space model evaluation of predictor and filter is simple. But this is a very restrictive class of models.

For general models we need to evaluate:

$$p(x_t|\underline{Y}_{t-1}) = \int p(x_{t-1}|\underline{Y}_{t-1}) p(x_t|x_{t-1}) dx_{t-1}.$$

Since $p(x_{t-1}|\underline{Y}_{t-1})$ and $p(x_t|x_{t-1})$ are known, one option is to use a numerical integration method (Kitagawa, 1988).

But computationally very intensive if x_t and y_t have a large dimension (typically more than 4).

Overview

- Estimation of the predictor and the filter distributions:
  - The Gaussian mixture model (valid for non-Gaussian linear models).
  - Sequential Monte Carlo method (also known as Particle filters).
- Estimation of the parameters (for example the F_t and G_t matrices in the linear model):
  - Self organising state-space models - Bayesian method on the parameters (Kitagawa, 1998).
  - A direct Particle filter approach (Huerzler and Kuensch, 2001).

Estimation using Gaussian mixture models

The linear model:

$$\begin{aligned} x_t &= F_t x_{t-1} + G_t w_t \\ y_t &= H_t x_t + e_t \end{aligned}$$

The errors w_t and e_t are not from a Gaussian distribution.

We approximate the error distributions with a Gaussian mixture model:

$$\begin{aligned} p_w(w_t) &\approx \sum_{j=1}^{K_w} \alpha_j \phi_j^w(w_t) \quad \phi_j^w \sim N(0, Q_j) \\ p_e(e_t) &\approx \sum_{i=1}^{K_e} \beta_i \phi_i^e(w_t) \quad \phi_i^e \sim N(0, \sigma_i^2). \end{aligned}$$

Therefore

$$p(x_t|x_{t-1}) \approx \sum_{j=1}^{K_w} \alpha_j \phi_j^w\{G_t^{-1}(x_t - F_t x_{t-1})\}, \qquad p(y_t|x_t) \approx \sum_{i=1}^{K_e} \beta_i \phi_j^e\{y_t - H_t x_t\}.$$

Suppose we have

$$p(x_{t-1}|\underline{Y}_{t-1}) = \sum_{s=1}^{M_{t-1}} \gamma_{t-1,s} \phi_{t-1,s}(x_{t-1}|\underline{Y}_{t-1})$$

where $\phi_{t-1,s}(x_{t-1}|\underline{Y}_{t-1}) \sim N(x_{t-1|t-1}^s, V_{t-1|t-1}^s)$.

Our object is to estimate $p(x_t|\underline{Y}_{t-1})$ and $p(x_t|\underline{Y}_t)$.

The predictor and filter estimates

The *predictor*:

$$p(x_t|\underline{Y}_{t-1}) = \int p(x_{t-1}|\underline{Y}_{t-1}) p(x_t|x_{t-1}) dx_{t-1}$$

$$= \sum_{s=1}^{M_{t-1}} \sum_{j=1}^{K_w} \alpha_j \gamma_{t-1,s} \underbrace{\int \phi_j^w\{G_t^{-1}(x_t - F_t x_{t-1})\} \phi_{t-1,s}(x_{t-1}|\underline{Y}_{t-1}) dx_{t-1}}_{\text{usual Gaussian predictor distribution}}$$

$$= \sum_{s=1}^{M_{t-1}} \sum_{j=1}^{K_w} \alpha_j \gamma_{t-1,s} \phi_{t,(s,j)}(x_t|\underline{Y}_{t-1}),$$

where $\phi_{t,(s,j)}(x_t|\underline{Y}_{t-1}) \sim N(x_{t|t-1}^{(s,j)}|V_{t|t-1}^{(s,j)})$ with

$$x_{t|t-1}^{(s,j)} = F_t x_{t-1|t-1}^s \text{ and } V_{t|t-1}^{(s,j)} = F_t V_{t-1|t-1}^s F_t^T + G_t Q_j G_t^T.$$

Relabel the sum limits ($L_t = K_w M_{t-1}$):

$$p(x_t|\underline{Y}_{t-1}) \;=\; \sum_{s=1}^{L_t} \delta_{t,s}\phi_{t,s}(x_t|\underline{Y}_{t-1}),$$

- The *filter*:

$$\begin{aligned} p(x_t|\underline{Y}_t) &= \frac{p(y_t|x_t)p(x_t|\underline{Y}_{t-1})}{p(y_t|\underline{Y}_{t-1})} \propto \sum_{s=1}^{L_t}\sum_{i=1}^{K_e} \delta_{t,s}\beta_i\phi_{t,s}(x_t|\underline{Y}_{t-1})\phi_i^e(Y_t|x_t) \\ &= \sum_{s=1}^{M_t} \gamma_{t,s}\phi_{t,s}(x_t|\underline{Y}_t), \end{aligned}$$

The mean and variance of $\phi_{t,s}(x_t|\underline{Y}_t)$ are evaluated using the usual Kalman filter equations.

Aside (evaluation of the filter components)

$$\begin{aligned} K_t^{s,i} &= V_{t|t-1}^s H_t^T \left(H_t V_{t|t-1}^s H_t^T + \sigma_i^2\right)^{-1} \quad \text{(the Kalman Factor)} \\ x_{t|t}^{s,i} &= x_{t|t-1}^s + K_t^{s,i}(y_t - H_t x_{t|t-1}^s) \\ V_{t|t}^{s,i} &= (I - K_t^{s,i} H_t) V_{t|t-1}^s \end{aligned}$$

$$\phi_{t,(s,i)}(x_t|\underline{Y}_t) \sim N(x_{t|t}^{s,i}, V_{t|t}^{s,i}).$$

- The *Smoother* can also be evaluated for the Gaussian mixture model. But this is a little bit more complicated (see Kitagawa, 1996).

Reducing the number of components

- For the distributions $p(x_t|\underline{Y}_{t-1})$ and $p(x_t|\underline{Y}_t)$ as t grows the number of Gaussian components is increasing exponentially!

- **A Solution** Trim the terms. Fix the number of Gaussian terms in a distribution. Choose the trimmed distributions which best fits $p(x_t|\underline{Y}_{t-1})$ and $p(x_t|\underline{Y}_t)$.

- Measure the distribution of best fit: use the Kullback-Leibler number:

  $$\int \log \frac{p(x_t|\underline{Y}_{t-1})}{q_L(x_t|\underline{Y}_{t-1})} p(x_t|\underline{Y}_{t-1}) dx_t,$$

  where $q_L(x_t|\underline{Y}_{t-1})$ is the trimmed distribution and includes only L Gaussian components.

- Choose the $q_L(x_t|\underline{Y}_{t-1})$ which minimises the Kullback-Leibler number (see Harrison and Stevens (1976)).

- There exist other computationally less intensive methods.

Sequential Monte Carlo methods

- Recall if $X_1, \ldots, X_M$ is an iid sample from the distribution P. Then an estimator of the distribution function is the empirical distribution

$$P_M(x) = \frac{1}{M} \sum_{j=1}^{M} I(x \leq X_j),$$

where I is the step function.

- This is the principle behind sequential Monte Carlo methods (particle filters). We approximate each distribution function by many realisations from that distribution. We then evaluate the probability of these realisations under a different distribution, these probabilities we use to

evaluate the new empirical distribution. At each time we will have the sample:

$$\begin{aligned} \{s_t^1, \ldots, s_t^M\} &\sim p(x_t | \underline{Y}_{t-1}) \\ \{u_t^1, \ldots, u_t^M\} &\sim p(x_t | \underline{Y}_t) \end{aligned}$$

The model

Recall the general nonlinear, non-Gaussian model:

$$\begin{aligned} y_t &= h_t(x_t, e_t) \qquad p_e(e) \text{ known} \\ x_t &= f_t(x_{t-1}, w_t) \quad p_w(w) \text{ known.} \end{aligned}$$

Assume that h is invertible, in the sense there exists a g, such that $e_t = g(x_t, y_t)$.

- We start at time $(t-1)$, and we assume $p_e(e)$ (equivalently $p(y_t|x_t)$), $p_w(w)$ (equivalently $p(x_t|x_{t-1})$) and $p(x_{t-1}|\underline{Y}_{t-1})$, where

$$p(x_{t-1}|\underline{Y}_{t-1}) = \frac{1}{M} \sum_{j=1}^{M} P(x \leq s_{t-1}^j)$$

are known.

- Our object is to estimate $p(x_t|\underline{Y}_{t-1})$ and $p(x_t|\underline{Y}_t)$ using samples from the above distributions.

The sequential Monte Carlo algorithm

Let M be large (this is our sampling size).

- **Step 1** The sampling
  - Sample M iid particles from $p(x_{t-1}|\underline{Y}_{t-1})$, this gives us the pool $\{s_{t-1}^1, \ldots, s_{t-1}^M\}$.
  - Sample M iid particles from $p_w(w)$, this gives us the pool $\{w_{t-1}^1, \ldots, w_{t-1}^M\}$.
  - Create new particles by combining s_{t-1}^j and w_t^j: $u_t^j = f_t(s_{t-1}^j, w_t^j)$. (equivalently sample $u_t^j \sim p(x_t|s_{t-1}^j)$).
    Since s_{t-1}^j was sampled from $p(x_{t-1}|\underline{Y}_{t-1})$, we have that $\{u_t^1, \ldots, u_t^M\}$ are samples from $p(x_t|\underline{Y}_{t-1})$.

- **Step 2** Since $\{u_t^1, \ldots, u_t^M\}$ are independent samples from $p(x_t|\underline{Y}_t)$, the empirical distribution estimator is:

$$p(x_t|\underline{Y}_{t-1}) \approx \frac{1}{M}\sum_{j=1}^{M} I(x_t \leq u_t^j)$$

- **Step 3** Estimation of $p(x_t|\underline{Y}_t)$
  - Use

$$\begin{aligned} p(x_t|\underline{Y}_t) &= \frac{p(y_t|x_t)p(x_t|\underline{Y}_{t-1})}{p(y_t|\underline{Y}_{t-1})} \\ p(x_t = u_t^j|\underline{Y}_t) &= \frac{p(y_t|x_t = u_t^j)p(x_t = u_t^j|\underline{Y}_{t-1})}{\sum_{i=1}^{M} p(y_t|x_t = u_t^i)p(x_t = u_t^i|\underline{Y}_{t-1})} \end{aligned}$$

(Note that $p(y_t|x_t = u_t^j)$ is the density function and $p(x_t = u_t^j|\underline{Y}_{t-1})$ the probability).

- Thus we need to estimate $p(y_t|x_t = u_t^i)$ and $p(x_t = u_t^i|\underline{Y}_{t-1})$:
-

$$p(y_t|x_t = u_t^i) = \frac{\partial g}{\partial y}\rfloor_{(u_t^j, y_t)} p_e\{g(u_t^j, y_t)\}$$

(do not worry about the form, just remember it can be evaluated). Let $\alpha_{t,i} = p(y_t|x_t = u_t^i)$.

- $p(x_t = u_t^i|\underline{Y}_{t-1}) = \frac{1}{M}$.
- Subsituting it in the above:

$$p(x_t = u_t^j|\underline{Y}_t) = \frac{p(y_t|x_t = u_t^j)p(x_t = u_t^j|\underline{Y}_{t-1})}{\sum_{i=1}^{M} p(y_t|x_t = u_t^i)p(x_t = u_t^i|\underline{Y}_{t-1})} = \frac{\alpha_{t,j}\frac{1}{M}}{\sum_{i=1}^{M}\alpha_{t,i}\frac{1}{M}} = \tilde{\alpha}_{t,j}.$$

Hence we have that $p(x_t = u_t^j|\underline{Y}_t) = \tilde{\alpha}_{t,j}$. We can use

$$p(x_t|\underline{Y}_t) = \sum_{j=1}^{M} \tilde{\alpha}_{t,j} I(x_t \leq u_t^j).$$

as an estimator of distribution function, $p(x_t|\underline{Y}_t)$.

In summary: we make M independent samples from the distribution $p(x_t|\underline{Y}_{t-1})$, which gives the sample $\{u_t^1, \ldots, u_t^M\}$. By using the empirical distribution we have that $p(x_t = u_t^j|\underline{Y}_{t-1}) = 1/M$. Next, using these 'particles' $\{u_t^1, \ldots, u_t^M\}$ we evaluate the conditional probability $p(y_t|x_t = u_t^j)$ - which we use to evaluate $p(x_t = u_t^j|\underline{Y}_{t-1})$.

- **Step 4** Obtaining $p(x_t|\underline{Y}_t)$ in a nice form. Alternatively by resampling $\{u_t^1,\dots,u_t^M\}$ we can obtain a more convenient distribution function. Resample the particles $\{u_t^1,\dots,u_t^M\}$ M times with $P(u_t^j) = \tilde{\alpha}_{t,j}$ to get the new sample $\{s_t^1,\dots,s_t^M\}$:

$$\{u_t^1,\dots,u_t^M\} \quad \Rightarrow \quad \{s_t^1,\dots,s_t^M\}.$$

Use

$$p(x_t|\underline{Y}_t) = \frac{1}{M}\sum_{j=1}^{M} I(x_t \le s_t^j),$$

as estimator of the distribution function $p(x_t|\underline{Y}_t)$.

The resampling of $\{u_t^1,\dots,u_t^M\}$ is similar to the Bootstrap filter (see Gordon et al (1993)).

Importance sampling

- To obtain an estimator of $p(x_t|\underline{Y}_t)$ we used a sample from $p(x_t|\underline{Y}_{t-1})$. In this case $p(x_t|\underline{Y}_{t-1})$ is called the *importance distribution*.

- It is possible to estimate $p(x_t|\underline{Y}_t)$ by sampling from another importance distribution, so long as the correct *importance weights* are used.

- Let $q(x_t|\underline{Y}_{t-1})$ be any other distribution. We take M iid particles from $q(x_t|\underline{Y}_{t-1})$: $\{\tilde{u}_t^1,\dots,\tilde{u}_t^M\}$

  (note that $q(x_t|\underline{Y}_{t-1})$ is now the importance distribution).

- We now evaluate $p(x_t = \tilde{u}_t^j|\underline{Y}_t)$

$$\begin{aligned} p(x_t = \tilde{u}_t^j|\underline{Y}_t) &= \frac{p(y_t|x_t = \tilde{u}_t^j)p(x_t = \tilde{u}_t^j|\underline{Y}_{t-1})}{\sum_{i=1}^M p(y_t|x_t = \tilde{u}_t^i)p(x_t = \tilde{u}_t^i|\underline{Y}_{t-1})} \\ &= \frac{p(y_t|x_t = \tilde{u}_t^j)\frac{p(x_t=\tilde{u}_t^j|\underline{Y}_{t-1})}{q(x_t=\tilde{u}_t^j|\underline{Y}_{t-1})}q(x_t = \tilde{u}_t^j|\underline{Y}_{t-1})}{\sum_{i=1}^M p(y_t|x_t = \tilde{u}_t^i)\frac{p(x_t=\tilde{u}_t^i|\underline{Y}_{t-1})}{q(x_t=\tilde{u}_t^i|\underline{Y}_{t-1})}q(x_t = \tilde{u}_t^i|\underline{Y}_{t-1})}. \end{aligned}$$

Since $q(x_t = \tilde{u}_t^i|\underline{Y}_{t-1})$ is the importance distribution, $\{\tilde{u}_t^j : j = 1,\dots,M\}$ is an independent sample from $q(x_t|\underline{Y}_{t-1})$. Therefore $q(x_t = \tilde{u}_t^i|\underline{Y}_{t-1}) = 1/M$. This gives

$$p(x_t = \tilde{u}_t^j|\underline{Y}_t) = \frac{p(y_t|x_t = \tilde{u}_t^j)w_{t,j}}{\sum_{i=1}^M p(y_t|x_t = \tilde{u}_t^i)w_{t,i}},$$

where $w_{t,i} = \frac{p(x_t = \tilde{u}_t^i | \underline{Y}_{t-1})}{q(x_t = \tilde{u}_t^i | \underline{Y}_{t-1})}$, and $w_{t,i}$ are called the importance weights.

- Now that we have $p(x_t = \tilde{u}_t^j | \underline{Y}_t)$ we can proceed using the rest of the sequential MC algorithm to obtain an estimator of the distribution $p(x_t | \underline{Y}_t)$.

- Importance sampling plays a role in estimation (amongst others).

Parameter Estimation

- Until now, we have assumed that the state-space functions f_t and g_t are completely known. In practice this will not be the case and many of the parameters need to be estimated.

- One method is to use the extended Kalman filters (Ljung, 1979). Sometimes the estimates are not satisfactory. This is because the method requires linearity. But even for linear models, when we want to estimate the parameters F_t, the state space model is nolonger linear.

The likelihood function:

$$p(\underline{Y}_T | \theta) = \prod_{t=1}^{T} p(y_t | \underline{Y}_{t-1}, \theta)$$

A likelihood aproach

- We note that

$$\begin{aligned} p(y_t | \underline{Y}_{t-1}, \theta) &= \int p(y_t | x_t, \theta) p(x_t | \underline{Y}_{t-1}, \theta) dx_{t-1} \\ &\approx \sum_{j=1}^{M} p(y_t | x_t = u_t^j, \theta) p(x_t = u_t^j | \underline{Y}_{t-1}, \theta) = \frac{1}{M} \sum_{j=1}^{M} \alpha_{t,j}(\theta), \end{aligned}$$

since $p(y_t | x_t = u_t^j, \theta) = \alpha_{t,j}(\theta)$ and $p(x_t = u_t^j | \underline{Y}_{t-1}, \theta) = 1/M$.

- Hence the log-likelihood is approximately

$$\log p(\underline{Y}_T | \theta) \approx \sum_{t=1}^{T} \log \left\{ \frac{1}{M} \sum_{j=1}^{M} \alpha_{t,j}(\theta) \right\}$$

- For every θ we use the sequential Monte Carlo method to estimate the likelihood $p(\underline{Y}_T | \theta)$. We choose the θ which maximises the likelihood $p(\underline{Y}_T | \theta)$ and use this as the parameter estimator.

- This methods requires sampling a new set of particles for each θ. This can be computationally very expensive. Also as each set of particles will be independent of the others - the results are quite irratic (likelihood function is not smooth).

- We now consider (i) a self organising state-space method (ii) a particle filter method, for the estimation of the parameters.

Self organising state-space models

Recall the nonlinear non-Gaussian state-space model:

$$
\begin{aligned}
y_t &= h(x_t, e_t) \\
x_t &= f(x_{t-1}, w_t).
\end{aligned}
$$

Suppose some parameters in the function f_t are unknown, call these unknown parameters θ. Rewrite:

$$x_t = f(x_{t-1}, w_t) = k(x_{t-1}, \theta, w_t).$$

Idea By placing a predictor distribution on θ, as well as for x_t we can also evaluate a predictor and filter for θ. That is we estimate:

$$p(x_t, \theta | \underline{Y}_{t-1}) \text{ and } p(x_t, \theta | \underline{Y}_t).$$

The method: self organising state-space models

- Let $z_t = (x_t, \theta)$. The new 'parameter included state space model' is:

$$
\begin{aligned}
z_t &= \begin{pmatrix} x_t \\ \theta \end{pmatrix} = \begin{pmatrix} k(x_{t-1}, \theta, w_t) \\ \theta \end{pmatrix} = \tilde{F}(x_{t-1}, \theta, w_t) \\
y_t &= \tilde{H}(z_t, e_t) = h(x_t, e_t).
\end{aligned}
$$

- Let θ have the initial distribution π_0. Since θ does not change over time, we do not need a transition probability $p(\theta_t | \theta_{t-1})$.

- Use the sequential MC method to estimate $p(x_t, \theta | \underline{Y}_{t-1})$ and $p(x_t, \theta | \underline{Y}_t)$ at each time point.

- To obtain the filter of only θ, integrate x_t:

$$p(\theta | \underline{Y}_t) = \int p(x_t, \theta | \underline{Y}_t) dx_t,$$

we have an estimate of the distribution of θ at each time point.

Example

Suppose y_t satisfies the model:

$$
\begin{aligned}
x_t &= F x_{t-1} + G w_t \\
y_t &= H x_t + e_t,
\end{aligned}
$$

where F is unknown. The new system equations is:

$$z_t = \begin{pmatrix} x_t \\ \theta \end{pmatrix} = \begin{pmatrix} F x_{t-1} + G w_t \\ F \end{pmatrix} = \tilde{F}(x_{t-1}, F, w_t)$$

Let π_0 be the initial distribution of F.

To start the sequential MC algorithm we make the pools:

$$\begin{array}{ccc} \{s_0^1,\ldots,s_0^M\} & \{F_0^1,\ldots,F_0^M\} & \{w_0^1,\ldots,w_0^M\} \\ \sim p_0(x) & \sim \pi_0(x) & \sim p_w(w) \end{array}$$

$$\{u_1^1,\ldots,u_1^M\}$$

where $u_1^j = \big((F_0^j s_0^j + w_0^j), F_0^j\big)$. Hence we can estimate the predictor. Resample as before to get the filter. And carry on the sequential MC algorithm...

Comments

- A small drawback is that since there is no transition density $p(F_t|F_{t-1})$, we will always have that $F_t^j \in \{F_0^1,\ldots,F_0^M\}$. Which can mean that the predictors and filters of F are not very informative.
- To overcome this problem, we can randomise F slightly;

  $$F_t = F_{t-1} + V_t, \qquad \mathrm{var} V_t = O(\frac{1}{t^2}).$$

  this means gives a more informative predictor and filter of F.
- The self-organsing state-space model can also be used to estimate time-varying parameters.

A filter sample method

This approach uses a likelihood method, but uses the idea of importance distributions to avoid resampling for different parameters θ.

The idea: Obtain one set of particles for a fixed θ_0: $\{\tilde{u}_t^1,\ldots,\tilde{u}_t^M : t=1,\ldots,T\}$ and $\{\tilde{s}_t^1,\ldots,\tilde{s}_t^M : t=1,\ldots,T\}$ and to use these particles to estimate the likelihood function for different θ.

- Recall that the estimator of the log-likelihood function is

  $$\log p(\underline{Y}_T|\theta) \approx \sum_{t=1}^{T} \log\Big\{\sum_{j=1}^{M} p(y_t|x_t = u_t^j,\theta)p(x_t = u_t^j|\underline{Y}_{t-1},\theta)\Big\}.$$

  Let $\theta=\theta_0$ and use sequential MC method to evaluate an estimator of the likelihood $p(\underline{Y}_T|\theta_0)$.

- This gives us the particles:

  $$\{\tilde{u}_t^1,\ldots,\tilde{u}_t^M : t=1,\ldots,T\} \text{ and } \{\tilde{s}_t^1,\ldots,\tilde{s}_t^M : t=1,\ldots,T\}.$$

- Using the particles $\{\tilde{u}_t^1,\ldots,\tilde{u}_t^M\}$ we have

  $$\begin{aligned} p(y_t|\underline{Y}_{t-1},\theta) &= \int p(y_t|x_t,\theta)p(x_t|\underline{Y}_{t-1},\theta)dx_{t-1} \\ &= \int p(y_t|x_t,\theta)\frac{p(x_t|\underline{Y}_{t-1},\theta)}{p(x_t|\underline{Y}_{t-1},\theta_0)}p(x_t|\underline{Y}_{t-1},\theta_0)dx_{t-1} \end{aligned}$$

  Now make M samples from the distribution $p(x_t|\underline{Y}_{t-1},\theta_0)$ (these are

$\{\tilde{u}_t^1, \ldots, \tilde{u}_t^M : t = 1, \ldots, T\})$ and we have

$$\begin{aligned} p(y_t|\underline{Y}_{t-1}, \theta) &\approx \frac{1}{M}\sum_{j=1}^{M} p(y_t|x_t = \tilde{u}_t^j, \theta)\frac{p(x_t = \tilde{u}_t^j|\underline{Y}_{t-1}, \theta)}{p(x_t = \tilde{u}_t^j|\underline{Y}_{t-1}, \theta_0)} \\ &= \frac{1}{M}\sum_{j=1}^{M} p(y_t|x_t = \tilde{u}_t^j, \theta) w_{t,j}(\theta_0, \theta), \end{aligned}$$

where

$$w_{t,j}(\theta_0, \theta) = \frac{p(x_t = \tilde{u}_t^j|\underline{Y}_{t-1}, \theta)}{p(x_t = \tilde{u}_t^j|\underline{Y}_{t-1}, \theta_0)},$$

$w_{t,j}(\theta_0, \theta)$ evaluated using a recursive equation.

Therefore the log likelihood is:

$$\log p(\underline{Y}_T|\theta) \approx \sum_{t=1}^{T} \log\left\{\frac{1}{M}\sum_{j=1}^{M} p(y_t|x_t = \tilde{u}_t^j, \theta) w_{t,j}(\theta_0, \theta)\right\}.$$

Use the θ which maximises $p(\underline{Y}_T|\theta)$ as an estimator of the true parameter.

- To summarise, for all θ the same particles were used for the estimation of the log-likelihood function $\log p(\underline{Y}_T|\theta)$. Just the probabilities and weights had to be recalculated. Analogous to the usual MLE, where for all θ the same set of observations are used to obtain the likelihood function. In this case, one can consider the particles as the observations (in some sense).
- Estimate more 'stable' than the resample likelihood method. Because

for different θ the same particles are used - gives a smoother likelihood $p(\underline{Y}_T|\theta)$.