V. Nonlinear state-space models

A. Extended Kalman filter

Linear state-space model:

State equation:

$$\underset{r\times 1}{\xi_{t+1}} = \underset{r\times r}{\mathbf{F}}\,\underset{r\times 1}{\xi_t} + \underset{r\times 1}{\mathbf{v}_{t+1}} \qquad \mathbf{v}_{t+1} \sim N(\mathbf{0},\mathbf{Q})$$

Observation equation:

$$\underset{n\times 1}{\mathbf{y}_t} = \underset{n\times k}{\mathbf{A}'}\,\underset{k\times 1}{\mathbf{x}_t} + \underset{n\times r}{\mathbf{H}'}\,\underset{r\times 1}{\xi_t} + \underset{n\times 1}{\mathbf{w}_t} \qquad \mathbf{w}_t \sim N(\mathbf{0},\mathbf{R})$$

Nonlinear state-space model:

State equation:

$$\underset{r\times 1}{\xi_{t+1}} = \underset{r\times 1}{\phi(\xi_t)} + \underset{r\times 1}{\mathbf{v}_{t+1}} \qquad \mathbf{v}_{t+1} \sim N(\mathbf{0},\mathbf{Q})$$

Observation equation:

$$\underset{n\times 1}{\mathbf{y}_t} = \underset{n\times 1}{\mathbf{a}(\mathbf{x}_t)} + \underset{n\times 1}{\mathbf{h}(\xi_t)} + \underset{n\times 1}{\mathbf{w}_t} \qquad \mathbf{w}_t \sim N(\mathbf{0},\mathbf{R})$$

Suppose at date t we have approximation to distribution of ξ_t conditional on

$$\Omega_t = \{\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_1, \mathbf{x}_t, \mathbf{x}_{t-1}, \ldots, \mathbf{x}_1\}$$

$$\xi_t | \Omega_t \sim N(\hat{\xi}_{t|t}, \mathbf{P}_{t|t})$$

goal: calculate $\hat{\xi}_{t+1|t+1}, \mathbf{P}_{t+1|t+1}$

State equation:

$$\xi_{t+1} = \phi(\xi_t) + \mathbf{v}_{t+1}$$

$$\phi(\xi_t) \simeq \phi_t + \Phi_t(\xi_t - \hat{\xi}_{t|t})$$

$$\underset{r\times 1}{\phi_t} = \phi(\hat{\xi}_{t|t})$$

$$\underset{r\times r}{\Phi_t} = \left.\frac{\partial \phi(\xi_t)}{\partial \xi_t'}\right|_{\xi_t = \hat{\xi}_{t|t}}$$

Forecast of state vector:

$$\xi_{t+1} = \phi_t + \Phi_t(\xi_t - \hat{\xi}_{t|t}) + \mathbf{v}_{t+1}$$

$$\hat{\xi}_{t+1|t} = \phi_t = \phi(\hat{\xi}_{t|t})$$

$$\mathbf{P}_{t+1|t} = \Phi_t \mathbf{P}_{t|t} \Phi_t' + \mathbf{Q}$$

Observation equation:

$$\mathbf{y}_t = \mathbf{a}(\mathbf{x}_t) + \mathbf{h}(\boldsymbol{\xi}_t) + \mathbf{w}_t$$

$$\mathbf{h}(\boldsymbol{\xi}_t) \simeq \mathbf{h}_t + \mathbf{H}_t'(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t-1})$$

$$\underset{n\times 1}{\mathbf{h}_t} = \mathbf{h}(\hat{\boldsymbol{\xi}}_{t|t-1})$$

$$\underset{n\times r}{\mathbf{H}_t'} = \left.\frac{\partial \mathbf{h}(\boldsymbol{\xi}_t)}{\partial \boldsymbol{\xi}_t'}\right|_{\boldsymbol{\xi}_t=\hat{\boldsymbol{\xi}}_{t|t-1}}$$

Note $\mathbf{x}_t$ is observed so no need to linearize $\mathbf{a}(\mathbf{x}_t)$

Approximating state equation:

$$\boldsymbol{\xi}_{t+1} = \boldsymbol{\phi}_t + \boldsymbol{\Phi}_t(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t}) + \mathbf{v}_{t+1}$$

Approximating observation equation:

$$\mathbf{y}_t = \mathbf{a}(\mathbf{x}_t) + \mathbf{h}_t + \mathbf{H}_t'(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t-1}) + \mathbf{w}_t$$

A state-space model with time-varying coefficients

Forecast of observation vector:

$$\mathbf{y}_{t+1} = \mathbf{a}(\mathbf{x}_{t+1}) + \mathbf{h}_{t+1} + \mathbf{H}_{t+1}'(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|t}) + \mathbf{w}_{t+1}$$

$$\hat{\mathbf{y}}_{t+1|t} = \mathbf{a}(\mathbf{x}_{t+1}) + \mathbf{h}_{t+1} = \mathbf{a}(\mathbf{x}_{t+1}) + \mathbf{h}(\hat{\boldsymbol{\xi}}_{t+1|t})$$

$$E(\mathbf{y}_{t+1} - \hat{\mathbf{y}}_{t+1|t})(\mathbf{y}_{t+1} - \hat{\mathbf{y}}_{t+1|t})' = \mathbf{H}_{t+1}'\mathbf{P}_{t+1|t}\mathbf{H}_{t+1} + \mathbf{R}$$

Updated inference:

$$\hat{\xi}_{t+1|t+1} = \hat{\xi}_{t+1|t} + \mathbf{K}_{t+1}(\mathbf{y}_{t+1} - \hat{\mathbf{y}}_{t+1|t})$$

$$\mathbf{K}_{t+1} = \mathbf{P}_{t+1|t}\mathbf{H}_{t+1}(\mathbf{H}'_{t+1}\mathbf{P}_{t+1|t}\mathbf{H}_{t+1} + \mathbf{R})^{-1}$$

Start from $\hat{\xi}_{0|0}$ and $\mathbf{P}_{0|0}$ reflecting prior information

Approximate log likelihood:

$$-\frac{Tn}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}\log|\mathbf{\Omega}_t| - \frac{1}{2}\sum_{t=1}^{T}\boldsymbol{\varepsilon}'_t\mathbf{\Omega}_t^{-1}\boldsymbol{\varepsilon}_t$$

$$\mathbf{\Omega}_t = \mathbf{H}'_t\mathbf{P}_{t|t-1}\mathbf{H}_t + \mathbf{R}$$

$$\boldsymbol{\varepsilon}_t = \mathbf{y}_t - \mathbf{a}(\mathbf{x}_t) - \mathbf{h}(\hat{\xi}_{t|t-1})$$

V. Nonlinear state-space models

A. Extended Kalman filter
B. Particle filter

State equation:

$$\underset{r\times 1}{\boldsymbol{\xi}_{t+1}} = \underset{r\times 1}{\boldsymbol{\phi}_t(\boldsymbol{\xi}_t, \mathbf{v}_{t+1})}$$

Observation equation:

$$\underset{n\times 1}{\mathbf{y}_t} = \underset{n\times 1}{\mathbf{h}_t(\boldsymbol{\xi}_t, \mathbf{w}_t)}$$

$\boldsymbol{\phi}_t(.)$ and $\mathbf{h}_t(.)$ known functions (may depend on unknown $\boldsymbol{\theta}$)

$\{\mathbf{w}_t, \mathbf{v}_t\}$ have known distribution (e.g., i.i.d., perhaps depend on $\boldsymbol{\theta}$)

$\Omega_t = \{\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_1\}$

$\Lambda_t = \{\boldsymbol{\xi}_t, \boldsymbol{\xi}_{t-1}, \ldots, \boldsymbol{\xi}_0\}$

output for step t:

$$p(\Lambda_t|\Omega_t)$$

represented by a series of particles:

$$\{\boldsymbol{\xi}_t^{(i)}, \boldsymbol{\xi}_{t-1}^{(i)}, \ldots, \boldsymbol{\xi}_0^{(i)}\}_{i=1}^{D}$$

Particle i is associated with weight $\hat{\omega}_t^{(i)}$ such that particles can be used to simulate draw from $p(\Lambda_t|\Omega_t)$, e.g. $E(\boldsymbol{\xi}_{t-1}|\Omega_t) = \sum_{i=1}^{D} \boldsymbol{\xi}_{t-1}^{(i)} \hat{\omega}_t^{(i)}$

Output of step $t+1$:

$p(\Lambda_{t+1}|\Omega_{t+1})$

keep particles $\{\xi_t^{(i)}, \xi_{t-1}^{(i)}, \ldots, \xi_0^{(i)}\}_{i=1}^{D}$

append $\{\xi_{t+1}^{(i)}\}_{i=1}^{D}$ and recalculate weights $\hat{\omega}_{t+1}^{(i)}$

and as byproduct we get an estimate of

$p(\mathbf{y}_{t+1}|\Omega_t)$

Method: Sequential Importance Sampling

At end of step t have generated

$\Lambda_t^{(i)} = \{\xi_t^{(i)}, \xi_{t-1}^{(i)}, \ldots, \xi_0^{(i)}\}$

from some known importance density

$g_t(\Lambda_t|\Omega_t) = \tilde{g}_t(\xi_t|\Lambda_{t-1}, \Omega_t) g_{t-1}(\Lambda_{t-1}|\Omega_{t-1})$

We will also have calculated (up to a constant that does not depend on ξ_t) the true value of $p_t(\Lambda_t|\Omega_t)$

so weight for particle i is proportional to

$\omega_t^{(i)} = \frac{p_t(\Lambda_t^{(i)}|\Omega_t)}{g_t(\Lambda_t^{(i)}|\Omega_t)}$

$$\omega_t^{(i)} = \frac{p_t(\Lambda_t^{(i)}|\Omega_t)}{g_t(\Lambda_t^{(i)}|\Omega_t)}$$

Step $t + 1$:

$$p_{t+1}(\Lambda_{t+1}|\Omega_{t+1}) = \frac{p(\mathbf{y}_{t+1}|\boldsymbol{\xi}_{t+1})p(\boldsymbol{\xi}_{t+1}|\boldsymbol{\xi}_t)p_t(\Lambda_t|\Omega_t)}{p(\mathbf{y}_{t+1}|\Omega_t)}$$

$$\propto \quad \underbrace{p(\mathbf{y}_{t+1}|\boldsymbol{\xi}_{t+1})}_{\text{known from obs eq}}\ \underbrace{p(\boldsymbol{\xi}_{t+1}|\boldsymbol{\xi}_t)}_{\text{known from state eq}}\ \underbrace{p_t(\Lambda_t|\Omega_t)}_{\text{known at } t}$$

$$\begin{aligned}
\omega_{t+1}^{(i)} &= \frac{p_{t+1}(\Lambda_{t+1}^{(i)}|\Omega_{t+1})}{g_{t+1}(\Lambda_{t+1}^{(i)}|\Omega_{t+1})} \\
&\propto \frac{p(\mathbf{y}_{t+1}|\boldsymbol{\xi}_{t+1}^{(i)})p(\boldsymbol{\xi}_{t+1}^{(i)}|\boldsymbol{\xi}_t^{(i)})p_t(\Lambda_t^{(i)}|\Omega_t)}{\tilde{g}_{t+1}(\boldsymbol{\xi}_{t+1}^{(i)}|\Lambda_t^{(i)},\Omega_{t+1})g_t(\Lambda_t^{(i)}|\Omega_t)} \\
&= \frac{p(\mathbf{y}_{t+1}|\boldsymbol{\xi}_{t+1}^{(i)})p(\boldsymbol{\xi}_{t+1}^{(i)}|\boldsymbol{\xi}_t^{(i)})}{\tilde{g}_{t+1}(\boldsymbol{\xi}_{t+1}^{(i)}|\Lambda_t^{(i)},\Omega_{t+1})} \frac{p_t(\Lambda_t^{(i)}|\Omega_t)}{g_t(\Lambda_t^{(i)}|\Omega_t)} \\
&= \tilde{\omega}_{t+1}^{(i)}\omega_t^{(i)}
\end{aligned}$$

$$\hat{\omega}_t^{(i)} = \frac{\omega_t^{(i)}}{\sum_{i=1}^{D} \omega_t^{(i)}}$$

$$\hat{E}(\boldsymbol{\xi}_{t-1}|\Omega_t) = \sum_{i=1}^{D} \hat{\omega}_t^{(i)}\boldsymbol{\xi}_{t-1}^{(i)}$$

$$\hat{P}(\xi_{1,t} > 0|\Omega_t) = \sum_{i=1}^{D} \hat{\omega}_t^{(i)}\delta_{[\xi_{1t}>0]}$$

$$\tilde{\omega}_{t+1}^{(i)} = \frac{p(\mathbf{y}_{t+1}|\xi_{t+1}^{(i)})p(\xi_{t+1}^{(i)}|\xi_t^{(i)})}{\tilde{g}_{t+1}(\xi_{t+1}^{(i)}|\Lambda_t^{(i)},\Omega_{t+1})}$$

$$\hat{p}(\mathbf{y}_{t+1}|\Omega_t) = \sum_{i=1}^{D} \tilde{\omega}_{t+1}^{(i)} \hat{\omega}_t^{(i)}$$

$$\hat{\mathcal{L}}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \log \hat{p}(\mathbf{y}_t|\Omega_{t-1})$$

Classical: choose $\boldsymbol{\theta}$ to max $\hat{\mathcal{L}}(\boldsymbol{\theta})$

Bayesian: draw $\boldsymbol{\theta}$ from posterior which is proportional to $p(\boldsymbol{\theta}) \exp[\hat{\mathcal{L}}(\boldsymbol{\theta})]$

How start algorithm for $t = 0$?

Draw $\xi_0^{(i)}$ from $p(\xi_0)$

(prior distribution or hypothesized unconditional distribution)

How choose importance density $\tilde{g}_{t+1}(\xi_{t+1}|\Lambda_t, \Omega_{t+1})$?

(1) Bootstrap filter

$\tilde{g}_{t+1}(\xi_{t+1}|\Lambda_t, \Omega_{t+1}) = p(\xi_{t+1}|\xi_t)$

known from state equation

$\xi_{t+1} = \boldsymbol{\phi}_t(\xi_t, \mathbf{v}_{t+1})$

But better performance from adaptive filters that also use $\mathbf{y}_{t+1}$

Note that for bootstrap filter

$$\tilde{g}_{t+1}(\xi_{t+1}|\Lambda_t, \Omega_{t+1}) = p(\xi_{t+1}|\xi_t)$$

$$\tilde{\omega}_{t+1}^{(i)} = \frac{p(\mathbf{y}_{t+1}|\xi_{t+1}^{(i)})p(\xi_{t+1}^{(i)}|\xi_t^{(i)})}{\tilde{g}_{t+1}(\xi_{t+1}^{(i)}|\Lambda_t^{(i)}, \Omega_{t+1})}$$

$$= p(\mathbf{y}_{t+1}|\xi_{t+1}^{(i)})$$

Separate problem for particle filter: one history $\Lambda_t^{(i)}$ comes to dominate the others ($\hat{\omega}_t^{(i)} \to 1$ for some i)

Partial solution to degeneracy problem:
Sequential Importance Sampling
with Resampling
Before finishing step t, now resample
$\{\Lambda_t^{(j)}\}_{j=1}^{D}$ with replacement
by drawing from the distribution

$$\Lambda_t^{(j)} = \begin{cases} \Lambda_t^{(1)} & \text{with probability } \hat{\omega}_t^{(1)} \\ \vdots & \\ \Lambda_t^{(D)} & \text{with probability } \hat{\omega}_t^{(D)} \end{cases}$$

Result: repopulate $\{\Lambda_t^{(j)}\}$ by replicating most likely elements (weights for $\Lambda_t^{(j)}$ are now $\hat{\omega}_t^{*(j)} = 1/D$).

(1) Resampling does not completely solve degeneracy because early-sample elements of $\Lambda_t^{(j)} = \{\xi_t^{(j)}, \xi_{t-1}^{(j)}, \ldots, \xi_0^{(j)}\}$ will tend to be the same for all j as t gets large

(2) Does help in the sense that have full set of particles to grow from t forward

(3) Have good inference about $p(\xi_{t-k}|\Omega_t)$ for small k

(4) Have poor inference about $p(\xi_{t-k}|\Omega_t)$ for large k

(separate smoothing algorithm can be used if goal is $p(\xi_t|\Omega_T)$)

Summary of bootstrap particle filter with resampling:

(1) Get initial set of D particles for date $t = 0$

(a) Set $\xi_{-100}^{(j)} = \mathbf{0}$ for $j = 1$

(b) Generate $\xi_t^{(j)} = \phi_0(\xi_{t-1}^{(j)}, \mathbf{v}_t^{(j)})$ for $t = -99, -98, \ldots, 0$

(c) Value of $\xi_0^{(j)}$ is one value for particle $j = 1$ for date 0

(d) repeat (a)-(c) for $j = 1, \ldots, D$ to populate $\{\xi_0^{(1)}, \xi_0^{(2)}, \ldots, \xi_0^{(D)}\}$

For any given $\boldsymbol{\theta}$ set $\ell_0(\boldsymbol{\theta}) = 0$ and for each $t = 1, 2, \ldots, T$ we then do the following:

(2) Compute $\tilde{\omega}_t^{(i)} = p(\mathbf{y}_t | \xi_t^{(i)})$ and update estimate of log likelihood:

$$\ell_t(\boldsymbol{\theta}) = \ell_{t-1}(\boldsymbol{\theta}) + \log\{D^{-1} \sum_{j=1}^{D} \tilde{\omega}_t^{(j)}\}$$

(3) Resample particles:

(a) Calculate $\hat{\omega}_t^{*(j)} = \tilde{\omega}_t^{(j)} / \left\{\sum_{j=1}^{D} \tilde{\omega}_t^{(j)}\right\}$

(b) Draw $u \sim U(0,1)$ and define $u^{(j)} = (u/D) + (j-1)/D$ for $j = 1, \ldots, D$.

(c) Find the indexes $i^1, \ldots, i^D$ such that $\sum_{k=1}^{i^j - 1} \hat{\omega}_t^{*(k)} < u^{(j)} \leq \sum_{k=1}^{i^j} \hat{\omega}_t^{*(k)}$

(4) Generate new particles:

Draw $\xi_{t+1}^{(j)}$ from $\phi_{t+1}(\xi_t^{ij}, \mathbf{v}_{t+1}^{(j)})$.

Repeat (2)-(4) for $t = 1,..,T$.

What do we do with estimate of log likelihood $\ell_T(\theta)$?

Best approach: embed within random-walk Metropolis-Hastings to generate draws of θ from posterior $p(\theta|\mathbf{Y})$ using prior $p(\theta)$.

(1) Generate initial draw $\theta^{(m)}$ for $m = 1$ and calculate $\ell_T(\theta^{(m)})$ and $p(\theta^{(m)})$.

(2) Generate $\tilde{\theta}^{(m+1)} \sim N(\theta^{(m)}, c\Lambda)$ and calculate $\ell_T(\tilde{\theta}^{(m+1)})$ and $p(\tilde{\theta}^{(m+1)})$.

(3) Set

$$\boldsymbol{\theta}^{(m+1)} = \begin{cases} \tilde{\boldsymbol{\theta}}^{(m+1)} & \text{with prob } \alpha \\ \boldsymbol{\theta}^{(m)} & \text{with prob } 1-\alpha \end{cases}$$

$$\alpha = \min\left\{ \frac{\ell_T(\tilde{\boldsymbol{\theta}}^{(m+1)})p(\tilde{\boldsymbol{\theta}}^{(m+1)})}{\ell_T(\boldsymbol{\theta}^{(m)})p(\boldsymbol{\theta}^{(m)})}, 1 \right\}.$$

Also possible to improve a lot on particle bootstrap by using better proposal density.

Example: use extended Kalman filter for proposal density in place of generating $\boldsymbol{\xi}_{t+1}^{(j)}$ from $\phi_t(\boldsymbol{\xi}_t^{(j)}, \mathbf{v}_{t+1})$.