PP3: Particle filters

BPF likelihood estimator

1a) We have $p(y_{1:n}) = \prod_{t=1}^{n} p(y_t | y_{1:t-1})$

so

$$\ell(y_{1:n}) = \log p(y_{1:n})$$
$$= \sum_{t=1}^{n} \log p(y_t | y_{1:t-1})$$

b) By marginalisation we have

$$p(y_t | y_{1:t-1}) = \int p(y_t, \alpha_t | y_{1:t-1}) \, d\alpha_t$$

$$= \int p(y_t | \alpha_t, y_{1:t-1}) \, p(\alpha_t | y_{1:t-1}) \, d\alpha_t$$

conditionally independent, given α_t

$$= \int g(y_t | \alpha_t) \, p(\alpha_t | y_{1:t-1}) \, d\alpha_t$$

c) If $\tilde{\alpha}_{t-1}^i \overset{\text{approx}}{\sim} p(\alpha_{t-1} | y_{1:t-1})$ and $\alpha_t^i \sim q(\alpha_t | \tilde{\alpha}_{t-1}^i)$, then the joint distribution of these samples is

$$(\alpha_t^i, \tilde{\alpha}_{t-1}^i) \overset{\text{approx}}{\sim} q(\alpha_t | \alpha_{t-1}) p(\alpha_{t-1} | y_{1:t-1})$$

$$= p(\alpha_{t-1}, \alpha_t | y_{1:t-1})$$

Hence, the marginal distribution of α_t^i is (approximately),

$$\boxed{p(\alpha_t | y_{1:t-1})}$$

d) From part c) we note that the particles $\{\alpha_t^i\}_{i=1}^N$ obtained after the simulation/propagation step of the bootstrap PF at time t can be viewed as an **unweighted** sample from $p(\alpha_t | y_{1:t-1})$

Thus, from part b, we can use these samples to approximate the one-step predictive pdf

$$p(y_t | y_{1:t-1}) = \int g(y_t | \alpha_t) p(\alpha_t | y_{1:t-1}) \, d\alpha_t$$

$$\approx \frac{1}{N} \sum_{i=1}^{N} g(y_t | \alpha_t^i) = \frac{1}{N} \sum_{i=1}^{N} \omega_t^i$$

where we used the fact that the weight is given by the observation likelihood in the bootstrap PF.

Finally, using a) we get the natural estimator

$$\ell(y_{1:n}) \approx \sum_{t=1}^{n} \log \left(\frac{1}{N} \sum_{i=1}^{N} \omega_t^i \right)$$

$$= \sum_{t=1}^{n} \left[\log \Omega_t - \log N \right]$$

"Log-trick" for numerical robustness,

a) The normalized weights are unaffected by the subtraction of a constant,

$$\frac{\tilde{\omega}_t^i}{\tilde{\Omega}_t} = \frac{e^{\log \omega_t^i - c_t}}{\sum_{j=1}^{N} e^{\log \omega_t^j - c_t}} = \frac{\cancel{e^{-c_t}}\, \omega_t^i}{\cancel{e^{-c_t}} \sum_{j=1}^{N} \omega_t^j}$$

b) Let I be the index of the maximum weight before rescaling,

$$I = \arg\max_i \{\log \omega_t^i\}$$

Hence, $c_t = \log \omega_t^I$

Since we subtract the same constant from all weights, the maximum rescaled log-weight is

$$\log \tilde{\omega}_t^I = \log \omega_t^I - c_t = 0$$

Hence all rescaled weights are bounded by

$$0 \leq \tilde{\omega}_t^i \leq \tilde{\omega}_t^I = \exp(0) = 1$$

It follows that

$$\tilde{\Omega}_t = \sum_{i=1}^{N} \tilde{\omega}_t^i \leq N$$

For the lower bound we use the fact that $\tilde{\omega}_t^I = 1$ and write

$$\tilde{\Omega}_t = \underbrace{\tilde{\omega}_t^I}_{=1} + \underbrace{\sum_{i \neq I} \tilde{\omega}_t^i}_{\geq 0} \geq 1$$

$$\therefore \quad 1 \leq \tilde{\Omega}_t \leq N$$

c) $$\log\left(\frac{1}{N}\sum_{i=1}^{N} w_t^i\right) = \log\left(\sum_{i=1}^{N} e^{\log w_t^i}\right) - \log N$$

$$= \log\left(\sum_{i=1}^{N} e^{\log \tilde{w}_t^i + c_t}\right) - \log N$$

$$= \log\left(e^{c_t}\sum_{i=1}^{N} \tilde{w}_t^i\right) - \log N$$

$$= \log(\tilde{\Omega}_t) + c_t - \log N$$

It follows that

$$l(y_{1:n}) \approx \sum_{t=1}^{n}\left[\log \tilde{\Omega}_t + c_t - \log N\right]$$

EM-algorithm

$$q(\alpha_t|\alpha_{t-1}) = \frac{1}{\sqrt{2\pi s^2}} \exp\left\{-\frac{1}{2}\frac{(\alpha_t - a\cdot\alpha_{t-1})^2}{s^2}\right\}$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\frac{\alpha_t^2 - 2a\overset{\alpha_t}{\alpha_{t-1}} + a^2\alpha_{t-1}^2}{s^2} - \frac{1}{2}\log s^2\right\}$$

We notice that α_t^2, $\alpha_t\alpha_{t-1}$, α_{t-1}^2 are the data and all multiplied with parameters so those are candidates for sufficient statistics

$$= \underbrace{\frac{1}{\sqrt{2\pi}}}_{h_q(\alpha_t,\alpha_{t-1})} \exp\left\{\underbrace{\begin{bmatrix}-\frac{1}{2s^2}\\ \frac{a}{s^2}\\ -\frac{a^2}{2s^2}\end{bmatrix}}_{\eta_q(\theta)}\cdot\underbrace{\begin{bmatrix}\alpha_t^2\\ \alpha_t\alpha_{t-1}\\ \alpha_{t-1}^2\end{bmatrix}}_{T_q(\alpha_t,\alpha_{t-1})} - \underbrace{\frac{1}{2}\log(s^2)}_{A_q(\theta)}\right\}$$

$$g(y_t|\alpha_t) = \frac{1}{\sqrt{2\pi b^2\exp\{-\alpha_t\}}} \exp\left\{-\frac{1}{2}\frac{y_t^2}{b^2\exp\{-\alpha_t\}}\right\}$$

$$= \frac{1}{\sqrt{2\pi\exp\{-\alpha_t\}}} \exp\left\{-\frac{1}{2}\frac{y_t^2\exp\{\alpha_t\}}{b^2} - \frac{1}{2}\log b^2\right\}$$

Same reasoning

$$= \underbrace{\frac{1}{\sqrt{2\pi\exp\{-\alpha_t\}}}}_{h_g(\alpha_t)} \exp\left\{\underbrace{\left[-\frac{1}{2b^2}\right]}_{\eta_g(\theta)}\cdot\underbrace{\left[y_t^2\exp\{\alpha_t\}\right]}_{T_g(\alpha_t)} - \underbrace{\frac{1}{2}\log b^2}_{A_g(\theta)}\right\}$$

Now we look at $Q(\theta, \tilde{\theta})$ skipping initial distribution

$$Q(\theta, \tilde{\theta}) = \sum_{t=2}^{n} \mathbb{E}[\log q(\alpha_t | \alpha_{t-1}) | y_{1:n}]$$

$$+ \sum_{t=1}^{n} \mathbb{E}[\log g(y_t | \alpha_t) | y_{1:n}]$$

$$= \ldots = \text{const} + \eta_q(\theta)\left(\sum_{t=2}^{n} \mathbb{E}[T_q(\alpha_t, \alpha_{t-1}) | y_{1:n}\right) - \frac{(n-1)}{2} \log s^2$$

$$+ \eta_g(\theta)\left(\sum_{t=1}^{n} \mathbb{E}[T_g(\alpha_t) | y_{1:n}]\right) - \frac{n}{2} \log b^2$$

let
$$t_1 = \sum_{t=2}^{n} \mathbb{E}[\alpha_t^2 | y_{1:n}]$$

$$t_2 = \sum_{t=2}^{n} \mathbb{E}[\alpha_t \alpha_{t-1} | y_{1:n}]$$

$$t_3 = \sum_{t=2}^{n} \mathbb{E}[\alpha_{t-1}^2 | y_{1:n}]$$

$$t_4 = \sum_{t=1}^{n} \mathbb{E}[y_t^2 \exp\{\alpha_t\} | y_{1:n}]$$

then

$$Q(\theta, \tilde{\theta}) = \text{const} - \frac{1}{2s^2} t_1 + \frac{a}{s^2} t_2 - \frac{a^2}{2s^2} t_3 - \frac{1}{2b^2} t_4 - \frac{(n-1)}{2} \log s^2 - \frac{n}{2} \log b^2$$

$$\frac{\partial Q}{\partial a} = \frac{1}{s^2} t_2 - \frac{a}{s^2} t_3 = 0 \Rightarrow a = \frac{t_2}{t_3}$$

$$\frac{\partial Q}{\partial s^2} = \frac{1}{2s^4} t_1 - \frac{a}{s^4} t_2 + \frac{a^2}{2s^4} t_3 - \frac{(n-1)}{2} \cdot \frac{1}{s^2} = 0 \Rightarrow s^2 = \frac{1}{n-1}\left(t_1 - \frac{t_2^2}{t_3}\right)$$

$$\frac{\partial Q}{\partial b^2} = \frac{1}{2b^4} t_4 - \frac{n}{2} \cdot \frac{1}{b^2} = 0 \Rightarrow b^2 = \frac{t_4}{n}$$