TSSL: Exercise session 3

The BPF likelihood estimator

In the particle filter the weighted particles $\{\alpha_t^i, \omega_t^i\}_{i=1}^N$ can be viewed as a *weighted sample* from the filtering distribution $p(\alpha_t \mid y_{1:t})$, in the sense that expected values w.r.t. this distribution can be estimated as

$$\int h(\alpha_t) p(\alpha_t \mid y_{1:t}) d\alpha_t \approx \Omega_t^{-1} \sum_{i=1}^N \omega_t^i h(\alpha_t^i)$$

for some function h. Similarly, the particles that we obtain *directly after the resampling step* can be viewed as an approximate sample from the same distribution. This is, if we denote the resampled particles by $\{\tilde{\alpha}_t^i\}_{i=1}^N$, then

$$\int h(\alpha_t) p(\alpha_t \mid y_{1:t}) d\alpha_t \approx \frac{1}{N} \sum_{i=1}^N h(\tilde{\alpha}_t^i)$$

is an alternative estimator of the filter expectation. In other words, the resampling can be thought of as turning the weighted sample into an unweighted one, replacing the importance weights with a random multiplicity of each particle.

1. In this exercise we will use the observation above to derive the likelihood estimator of the bootstrap particle filter.

   (a) Write the log-likelihood $\ell(y_{1:n}) := \log p(y_{1:n})$ in terms of the one-step predictive pdfs $p(y_t \mid y_{1:t-1})$.

   (b) Express $p(y_t \mid y_{1:t-1})$ in terms of the observation model $g(y_t \mid \alpha_t)$ and the state predictive distribution $p(\alpha_t \mid y_{1:t-1})$.

   Hint: Use the Markov property of the model and a marginalization over α_t.

   (c) If $\tilde{\alpha}_{t-1}^i$ is an (approximate) sample from the filtering distribution at time $t-1$, $p(\alpha_{t-1} \mid y_{1:t-1})$, then what can be said about the distribution of $\alpha_t^i \sim q(\alpha_t \mid \tilde{\alpha}_{t-1}^i)$ obtained after the *propagation step* of the bootstrap particle filter?

   (d) Putting the above together, derive an estimator of the log-likelihood $\ell(y_{1:n})$.

Taking the Exponent of Log Weights

When implementing the particle filter in practice it is a good idea to work with the logarithm of the importance weights instead of the weights directly. This is for reasons of numerical stability; a very small weight will be a large negative number instead when looking at the log-weight. A problem that arises when going back to the original weights is that these large negative numbers becomes numerically 0 when taking the exponential. If this happens simultaneously for *all* weights, the normalization constant $\Omega_t = \sum_{i=1}^N \omega_t^i$ will also be (numerically) equal to zero, resulting in an undefined expression $0/0$ when normalizing the weights.

To combat this we need to do some scaling of the log-weights. The way we do it is that we let $c_t = \max_{i\in\{1,2,\dots,N\}} \log \omega_t^i$ and subtract this value from each of the log-weights. That is, we compute that variables

$$\log \tilde{\omega}_t^i = \log \omega_t^i - c_t.$$

1. In this exercise we will look at what this means for the normalized weights and the log-likelihood estimate.

   (a) What happens if we normalize the scaled $\tilde{\omega}_t^i$ weights?
   Calculate $\tilde{\Omega}_t = \sum_{i=1}^N \tilde{\omega}_t^i$ and look at $\frac{\tilde{\omega}_t^i}{\tilde{\Omega}_t}$. How does this relate to $\frac{\omega_t^i}{\Omega_t}$?

   (b) What is the maximum value of the rescaled weights, $\max_i \tilde{\omega}_t^i$? Using this, compute an upper and a lower bound for $\tilde{\Omega}_t$. *(These bounds show that we get a numerically well-behaved computation when normalizing the weights by $\tilde{\Omega}_t$.)*

   (c) For the log-likelihood estimate using the bootstrap particle filter we need to compute $\log(\frac{1}{N}\sum_{i=1}^N \omega_t^i)$. Express this in terms of the sum of the re-scaled weights $\tilde{\Omega}_t$ and the constant c_t.

Sufficient Statistics and EM Algorithm

For the EM-algorithm we noticed in the lecture that if the model belongs to the exponential family performing the E-step reduced down to calculating the smoothed sum of sufficient statistics, while the M-step reduced to solving an equation.

We look at the Stochastic Volatility model given by the equations:

$$\begin{cases} \alpha_t = a\alpha_{t-1} + s\eta_t, & \eta_t \sim \mathcal{N}(0,1), \\ y_t = b\exp(\alpha_t/2)\varepsilon_t, & \varepsilon_t \sim \mathcal{N}(0,1). \end{cases}$$

Here the parameters are $\theta = (a, s^2, b^2)$.

- Find $\mathbf{T}(\alpha_{1:t}, y_{1:t})$, $\mathbf{n}(\theta)$ and $A(\theta)$ for this model.

1. Look at $q(\alpha_t \,|\, \alpha_{t-1})$ and find the corresponding functions.
2. Look at $g(y_t \,|\, \alpha_t)$ and find the corresponding functions.
3. Finally take the sum $\sum_{t=2}^{n} \log q(\alpha_t \,|\, \alpha_{t-1}) + \sum_{t=1}^{n} g(y_t \,|\, \alpha_t)$ and use the previous parts to write this on the form $\mathbf{T}(\alpha_{1:t}, y_{1:t}) \cdot \mathbf{n}(\theta) - A(\theta) +$ const.

- Find the parameter updating equation that maximizes $\mathbf{T}(\alpha_{1:t}, y_{1:t}) \cdot \mathbf{n}(\theta) - A(\theta)$.

  1. Take the derivative with respect to the three parameters.
  2. Solve the derivatives equal to zero *simultaneously*.