1 Updating the VB Approximation

An issue with using the Particle Filter from $T + 1$ to $T + S$ is that it only gives you an estimate of $p(y_{T+1:T+S}|\theta)$, and not $p(y_{1:T+S}|\theta)$. This could be an issue as the VB algorithm requires the distribution

$$\ln(p(y_{1:T+S}, \theta, x_{T+S+1})) = \ln(p(\theta)) + \ln(p(y_{1:T+S}|\theta)) + \ln(p(x_{T+S+1}|y_{1:T+S}, \theta)). \tag{1}$$

as part of the ELBO,

$$\mathcal{L}(\boldsymbol{\lambda}) = \int q_{\boldsymbol{\lambda}}(\boldsymbol{\theta}, x_{T+S+1}|y_{1:T+S}) \ln\left(\frac{p(y_{1:T+S}, \boldsymbol{\theta}, x_{T+S+1})}{q_{\boldsymbol{\lambda}}(\boldsymbol{\theta}, x_{T+S+1}|y_{1:T+S})}\right) d\boldsymbol{\theta} dx_{1:T}. \tag{2}$$

This issue should be avoidable but requires a slightly different derivation of the ELBO. Note that

$$
\begin{aligned}
\ln(p(y_{T+1:T+S}|y_{1:T})) &= \int_{\theta, x_{T+S+1}} q(\theta, x_{T+1}|y_{1:T+S}) \ln(p(y_{T+1:T+S}|y_{1:T})) d_\theta d_{x_{T+S+1}} \\
&= \int_{\theta, x_{T+S+1}} q(\theta, x_{T+S+1}|y_{1:T+S}) \ln\left(\frac{p(y_{T+1:T+S}, \theta, x_{T+1}|y_{1:T})}{p(x_{T+1}, \theta|y_{1:T+S})}\right) d_\theta d_{x_{T+S+1}} \\
&= \int_{\theta, x_{T+S+1}} q(\theta, x_{T+S+1}|y_{1:T+S}) \\
&\times \ln\left(\frac{p(y_{T+1:T+S}, \theta, x_{T+1}|y_{1:T})q(\theta, x_{T+S+1}|y_{1:T+S})}{p(x_{T+1}, \theta|y_{1:T+S})q(\theta, x_{T+S+1}|y_{1:T+S})}\right) d_\theta d_{x_{T+1}} \\
&= \int_{\theta, x_{T+S+1}} q(\theta, x_{T+S+1}|y_{1:T+S}) \ln\left(\frac{p(y_{T+1:T+S}, \theta, x_{T+1}|y_{1:T})}{q(\theta, x_{T+S+1}|y_{1:T+S})}\right) d_\theta d_{x_{T+1}}
\end{aligned}
\tag{3}
$$

$$+ \int_{\theta, x_{T+S+1}} q(\theta, x_{T+S+1}|y_{1:T+S}) \ln\left(\frac{q(\theta, x_{T+S+1}|y_{1:T+S})}{p(\theta, x_{T+1}|y_{1:T+S})}\right) d_\theta d_{x_{T+1}} \tag{4}$$

$$= \mathcal{L}_2(\boldsymbol{\lambda}) + KL[q(\theta, x_{T+S+1}|y_{1:T+S})||p(\theta, x_{T+S+1}|y_{1:T+S})].$$

As (4) is the Kullback-Leibler divergence from q to p, it is minimised by maximising either (2) or (3), ie. maximising an ELBO that contains $\ln(p(y_{T+1:T+S}, \theta, x_{T+S+1}|y_{1:T}))$ instead of $\ln(p(y_{1:T+S}, \theta, x_{T+S+1}))$ is still valid. $\mathcal{L}_2(\boldsymbol{\lambda})$, given by (3), contains

$$\ln(p(y_{T+1:T+S}, \theta, x_{T+S+1}|y_{1:T})) = \ln(p(\theta|y_{1:T})) + \ln(p(y_{T+1:T+S}|\theta, y_{1:T})) + \ln(p(x_{T+S+1}|y_{1:T+S}, \theta)). \tag{5}$$

This introduces a few problems,

1. We do not have a parametric form of the marginal posterior distribution $p(\theta|y_{1:T})$. We do, however, have a parametric approximation available from the earlier VB: $q(\theta|y_{1:T})$. This can be feasibly used, but we would have to be comfortable with changing the model specification.

2. The particle filter estimates must be conditioned on $y_{1:T}$.

2 VB Particle Filtering

We want samples from distribution $p(x_{T+1:T+S}|\theta, y_{1:T+S})$, which we decompose as

$$p(x_{T+1:T+S}|\theta, y_{1:T+S}) = p(x_{T+1}|y_{1:T}, \theta)p(y_{T+1}|x_{T+1}, \theta) \prod_{t=T+2}^{T+S} (p(y_t|x_t, \theta)p(x_t|x_{t-1}, \theta)) \quad (6)$$

However, this is unavailable so instead we draw particles $x_{T+1:T+S}^{(1:N)}$ from a proposal distribution $f(x_{T+1:T+S})$, and then use importance sampling weights to correct the approximation. The weights for particle vector i are given by

$$w^{(i)} \propto \frac{p(x_{T+1:T+S}^{(i)}|\theta, y_{1:T+S}, \theta)}{f(x_{T+1:T+S}^{(i)})}. \quad (7)$$

A particle filter operates by positing a proposal distribution that factorises according to $f(x_{T+1:T+S}) = \prod_{t=T+1}^{T+S} f(x_t)$, and drawing samples and calculating the weights for each time period sequentially. In this case, the weights for particles $x_{T+1}^{(i)}$ from the proposal distribution $f(x_{T+1})$ are given by

$$w_{T+1}^{(i)} \propto \frac{p(x_{T+1}^{(i)}|y_{1:T}, \theta)p(y_{T+1}|x_{T+1}^{(i)}, \theta)}{f(x_{T+1}^{(i)})} \quad (8)$$

and weights at a future time t are given by the recursion

$$w_t^{(i)} \propto w_{t-1}^{(i)} \frac{p(y_t|x_t^{(i)}, \theta)p(x_t^{(i)}|x_{t-1}^{(i)}, \theta)}{f(x_t^{(i)})}. \quad (9)$$

Equation (8) is problematic, we cannot easily evaluate the distribution $p(x_{T+1}|y_{1:T}, \theta)$. As before, we can approximate this with the distribution $q(x_{T+1}|\theta, y_{1:T})$. If this is satisfactory, choose $f(x_{T+1})$ to be $q(x_{T+1}|\theta, y_{1:T})$ and each other $f(x_t)$ to be $p(x_t|x_{t-1}, \theta)$. If the particles are resampled at each step, the weights reduce to

$$w_t^i \propto p(y_t|x_t^i, \theta) \quad (10)$$

for all t.

3 Discussion

Implementing particle filter updates introduces numerous problems, each of which seems to be solvable if we are willing to use the existing approximation of $q(\theta, x_{T+1}|y_{1:T})$ in place of the true posterior $p(\theta, x_{T+1}|y_{1:T})$. This implicitly changes the model prior used in an extremely informative way and not particularily desirable way - the prior must have the exact form that results in this particular multivariate normal posterior. If we accept the original VB approximation as good enough then maybe this won't be such a big deal; x_{T+S+1} is not included in the above substitution though there likely some indirect impact.