Exercise: Particle Filter

Consider the following stochastic linear dynamical system from last week's exercise:

$$y_t = x_t^{(1)} + x_t^{(2)} + r_t$$

$$x_t^{(1)} = q_t^{(1)}$$

$$x_t^{(2)} = \gamma x_{t-1}^{(2)} + \alpha W_{t-1} x_{t-1}^{(1)} + q_t^{(2)}$$

$$W_{t-1} = \begin{cases} 0 & y_{t-1} \geq \theta \\ -1 & y_{t-1} < \theta \end{cases}$$

with parameters $\gamma < 1$ and $m_0 = 0$. $q_t^{(i)} \in \mathcal{N}(0, Q^{(i)})$ and $r_t \in \mathcal{N}(0, R)$ are Gaussian distributed white noise (i.i.d.). As discussed in the lecture, this system describes the pitch trajectory of a bird during the reinforcement learning task where y_t is the mean-subtracted pitch of syllable rendition t, and where the hidden state $x_t^{(1)}$ determines the exploration and $x_t^{(2)}$ integrates the learned changes. The bird learns by correlating exploration and reward (whether or not it gets punished) with a learning rate α. W_{t-1} is zero if the bird escapes punishment (pitch y_{t-1} was above threshold θ) and negative if a white noise burst was delivered for punishment (pitch y_{t-1} was below threshold θ).

Compare Particle Filter with Kalman Filter

1. Insert the missing line in bootstrapParticleFilter.m that calculates the weights according to the Bootstrap algorithm and implement the resampling function 'sys_resampling' as seen in the lecture.
2. Use the solution from last week's exercise to estimate the hidden state sequence of the dynamical system above with the particle Filter. How does it compare to the Kalman filter estimation? Run the estimation again for the same generated data and explain why the Kalman filter estimation does not change, whereas the particle filter estimation does.
3. Run the particle filter 20 times and compare mean and standard deviation of the log-likelihood for different numbers of particles (ex. 50, 100, 500). What is the minimum number needed to get a reasonable estimate?
4. Plot the weights at time $t = 500$ and $t = 6500$ and describe their behavior for 50 and for 500 particles.

Reward prediction error

5. It is biologically better feasible to learn by corrolating exploration with the mean subtracted reward (reward prediction error) instead of reward by itself. The new state equations then is
$$x_t^{(2)} = \gamma x_{t-1}^{(2)} + \alpha(W_{t-1} - \overline{W}_{t-1}) x_{t-1}^{(1)} + q_t^{(2)}$$
Where $\overline{W}_t = (1-\beta)\overline{W}_{t-1} + \beta W_{t-1}$ is the mean subtracted reward with β being the discounting factor. This model is already implemented in the solution of last week's exercise. Simulate pitch reinforcement data using this model by setting the variable 'model' to 2 and adjust the learning rate α such that model bird learns to increase its pitch by 25 Hz.

6. Modify the state transition function in bootstrapParticleFilter.m such that you can estimate the hidden state using mean subtracted reward.

   Hint: there is no noise in the mean subtracted reward. Thus, it is enough to add another input to the state transition function 'state_transition' and keep track of the deterministic mean subtracted reward variable. As an alternative you can also re-compute the mean subtracted reward inside the state transition function using the inputs you already have available.

7. Why is it not possible to estimate the hidden state with the mean subtracted reward using a Kalman filter?