DS-GA 3001.001 Special Topics in Data Science: Modeling Time Series
Handout: Particle filtering

1 Motivation

In some situations the assumptions of LDS are too restrictive, for instance when we need a non-gaussian emission density (e.g. because observations are strictly positive, or discrete counts). In such cases we no longer have a simple closed form solution for the posterior marginals $\mathrm{P}(\mathbf{z}_i|\mathbf{x}_{1:t})$. Nonetheless, if we have a way to generate samples from this distribution, then we can use them approximate the desired posterior marginals and still be able to compute useful things (e.g. expectations for the E-step during maximum likelihood learning).

In general approximate inference via sampling allows us to compute expectations and other quantities of interest which are difficult to write down in closed form. In particular, if we have generated N samples $\mathbf{z}_i^{(n)}$, then any expectation of a function $f(\cdot)$ can be approximatively computed as an empirical mean of its values, $\mathbb{E}_{\mathrm{P}(\mathbf{z}_i)}[f(\mathbf{z}_i)] = \frac{1}{N}\sum_n f\left(\mathbf{z}_i^{(n)}\right)$ where $\mathbf{z}_i^{(n)} \sim \mathrm{P}(\mathbf{z}_i)$.

Particle filtering provides an efficient way to generate such samples for sequential latent observations. If for the linear gaussian version inference involved recursive equations for updating the parameters of the gaussian marginal posterior, here we are looking for recursive updates for the individual samples as they propagate through the sequence. This procedure involves 2 steps:[1] a) given a set of samples for $\mathbf{z}_i^{(n)} \sim \mathrm{P}(\mathbf{z}_i|\mathbf{x}_{1:i-1})$ how do we change these to incorporate $\mathbf{x}_i$; and b) given a set of samples from $\mathrm{P}(\mathbf{z}_i|\mathbf{x}_{1:i})$ how do we translate these into a new set of samples for the next latent state $\mathbf{z}_{i+1}$.

2 Incorporating observations

Assume we are given a set of N samples from $\mathbf{z}_i^{(n)} \sim \mathrm{P}(\mathbf{z}_i|\mathbf{x}_{1:i-1})$, and that we want to estimate $\mathbb{E}_{\mathbf{z}_i|\mathbf{x}_{1:i}}[f(\mathbf{z}_i)]$. Using Bayes rule and the conditional dependencies implied by the graphical model ($\mathrm{P}(\mathbf{x}_i|\mathbf{z_i}, \mathbf{x}_{1:i-1}) = \mathrm{P}(\mathbf{x}_i|\mathbf{z_i})$, same as for HMMs and LDS), we have:

$$\begin{aligned}
\mathbb{E}[f(\mathbf{z}_i)] &= \int f(\mathbf{z}_i)\mathrm{P}(\mathbf{z}_i|\mathbf{x}_{1:i})\, d\mathbf{z}_i && (1)\\
&= \int f(\mathbf{z}_i)\mathrm{P}(\mathbf{z}_i|\mathbf{x}_{1:i-1}, \mathbf{x}_i)\, d\mathbf{z}_i && (2)\\
&= \frac{\int f(\mathbf{z}_i)\mathrm{P}(\mathbf{x}_i|\mathbf{z}_i)\,\mathrm{P}(\mathbf{z}_i|\mathbf{x}_{1:i-1})\, d\mathbf{z}_i}{\int \mathrm{P}(\mathbf{x}_i|\mathbf{z}_i)\,\mathrm{P}(\mathbf{z}_i|\mathbf{x}_{1:i-1})\, d\mathbf{z}_i} && (3)
\end{aligned}$$

which can be approximated as:

$$\mathbb{E}[f(\mathbf{z}_i)] \approx \sum_n w_i^{(n)} f\left(\mathbf{z}_i^{(n)}\right), \text{where } w_i^{(n)} = \frac{\mathrm{P}\left(\mathbf{x}_i|\mathbf{z}_i^{(n)}\right)}{\sum_m \mathrm{P}\left(\mathbf{x}_i|\mathbf{z}_i^{(m)}\right)} \quad (4)$$

where the same samples are used in the nominator and denominator. This is a form of *importance sampling*, where we draw samples from the 'wrong' distribution $\mathrm{P}(\mathbf{z}_i|\mathbf{x}_{1:i-1})$, but then reweigh these samples to correctly compensate for the mismatch. The weights $w_i^{(n)}$ satisfy $0 \leq w_i^{(n)} \leq 1$ and $\sum_n w_i^{(n)} = 1$.

3 Predicting subsequent latent state

At this point we have generated samples $\mathbf{z}_i^{(n)}$ and their corresponding importance weights $w_i^{(n)}$. We want to use these to generate predictions for $\mathbf{z}_{i+1}$, as a set of samples drawn from the marginal posterior

[1] These steps are presented in the reverse order from the two steps of the derivation for the kalman filter updates, but they are essentially trying to achieve the same computations.

$\mathrm{P}\left(\mathbf{z}_{i+1}|\mathbf{x}_{1:i}\right)$. Marginalizing out the unknown previous state, we can write the distribution we are interested in as:

$$\mathrm{P}\left(\mathbf{z}_{i+1}|\mathbf{x}_{1:i}\right) = \int \mathrm{P}\left(\mathbf{z}_{i+1}|\mathbf{z}_i, \mathbf{x}_{1:i}\right) \mathrm{P}\left(\mathbf{z}_i|\mathbf{x}_{1:i}\right) d\mathbf{z}_i = \int \mathrm{P}\left(\mathbf{z}_{i+1}|\mathbf{z}_i\right) \mathrm{P}\left(\mathbf{z}_i|\mathbf{x}_{1:i}\right) d\mathbf{z}_i \tag{5}$$

$$\approx \sum_n w_i^{(n)} \mathrm{P}\left(\mathbf{z}_{i+1}|\mathbf{z}_i^{(n)}\right) \tag{6}$$

where we used the conditional independence relationship $\mathrm{P}\left(\mathbf{z}_{i+1}|\mathbf{z}_i, \mathbf{x}_{1:i}\right) = \mathrm{P}\left(\mathbf{z}_{i+1}|\mathbf{z}_i\right)$ and then noticed that the resulting expression has the same general form as derived in the previous step, with $f(\cdot)$ replaced by $\mathrm{P}\left(\mathbf{z}_{i+1}|\cdot\right)$. From the perspective of $\mathbf{z}_{i+1}$ the above expression is a mixture with N components; the probability of each component is $w_i^{(n)}$ and the corresponding observation model is $\mathrm{P}\left(\mathbf{z}_{i+1}|\mathbf{z}_i^{(n)}\right)$. To sample from this mixture model we draw N independent class assignments c_m $\mathbf{w}_i$, then for each draw one sample from $\mathbf{z}_{i+1}^{(m)} \sim \mathrm{P}\left(\mathbf{z}_{i+1}|\mathbf{z}_i^{(c_m)}\right)$. Step a is then repeated using the unweighted samples $\mathbf{z}_{i+1}^{(m)}$ as input, and so on.