## Prediction with mixture model

The task of prediction with a mixture has good sense only if both components and pointer model are dynamic. I.e. the components are of the predictive model type and the pointer model depends on the previous value of the pointer variable.

### Zero step prediction - example

The components are

$$f\left(y_t|y_{t-1},u_t,\theta\right) \text{ or } y_t=ay_{t-1}+bu_t+e_t$$

with the input $u$ fixed and given beforehand and $\theta=\{a,b\}$. Variance of $e$ is given.

Pointer model

$$f\left(c_t|c_{t-1},\alpha\right)=\alpha_{c_t|c_{t-1}}$$

With this the Bayes rule is ($d_t=\{y_t,u_t\}$ and under NCC)

$$f\left(c_t,c_{t-1},\theta,\alpha|d\left(t\right)\right)\propto f\left(y_t,c_t,c_{t-1},\theta,\alpha|d\left(t-1\right)\right)=$$

$$=\underbrace{f\left(y_t|y_{t-1},u_t,c_t,\theta\right)f\left(\theta|d\left(t-1\right)\right)}_{\text{updt of component}}\times\underbrace{f\left(c_t|c_{t-1},\alpha\right)f\left(\alpha|d\left(t-1\right)\right)}_{\text{updt of pointer models}}f\left(c_{t-1}|d\left(t-1\right)\right)$$

Classification

$$f\left(c_t|d\left(t\right)\right)\propto\sum_{c_{t-1}}\int_{\theta^*}\int_{\alpha^*}f\left(c_t,c_{t-1},\theta,\alpha|d\left(t\right)\right)d\alpha d\theta=$$

$$= \int_{\theta^*} f\left(y_t|y_{t-1}, u_t, c_t, \theta\right) f\left(\theta|d\left(t-1\right)\right) d\theta \times$$

$$\times \sum_{c_{t-1}} \left[ \int_{\alpha^*} f\left(c_t|c_{t-1}, \alpha\right) f\left(\alpha|d\left(t-1\right)\right) d\alpha \times f\left(c_{t-1}|d\left(t-1\right)\right) \right]$$

If we denote

$$f\left(c_t|d\left(t\right)\right) = w_t \rightarrow f\left(c_{t-1}|d\left(t-1\right)\right) = w_{t-1}$$

$$\int_{\theta^*} f\left(y_t|y_{t-1}, u_t, c_t, \theta\right) f\left(\theta|d\left(t-1\right)\right) d\theta = \hat{m}_t$$

$$\int_{\alpha^*} f\left(c_t|c_{t-1}, \alpha\right) f\left(\alpha|d\left(t-1\right)\right) d\alpha = \hat{\alpha}_{c_t|c_{t-1}}$$

we can write (in brief) evolution of the pointer estimate

$$w_t \propto \hat{m}_t \hat{\alpha}_{c_t|c_{t-1}} w_{t-1}$$

where $\hat{\alpha}_{c_t|c_{t-1}} w_{t-1}$ is a stationary prediction of the pointer (with no information of $y_t$) and $\hat{m}_t$ brings correction with respect to the measured $y_{t-1}$.

Prediction

$$f\left(y_t|d\left(t-1\right)\right) \propto \sum_{c_t} \sum_{c_{t-1}} \int_{\theta^*} \int_{\alpha^*} f\left(y_t, c_t, c_{t-1}, \alpha, \theta|d\left(t-1\right)\right) d\alpha d\theta =$$

$$= \sum_{c_t} \left\{ \int_{\theta^*} f\left(y_t|y_{t-1}, u_t, c_t, \theta\right) f\left(\theta|d\left(t-1\right)\right) d\theta \times \right.$$

$$\times \sum_{c_{t-1}} \left[ \int_{\alpha^*} f\left(c_t|c_{t-1},\alpha\right) f\left(\alpha|d\left(t-1\right)\right) d\alpha \times f\left(c_{t-1}|d\left(t-1\right)\right) \right] \Bigg\} =$$

$$\sum_{c_t} \hat{m}_t \hat{\alpha}_{c_t|c_{t-1}} w_{t-1}$$

which is weighted sum of predictions form components, where the weights are predictions of th pointer without use of $y_t$.

Remark

The difference between $\hat{m}_t$ in evaluation the weight $w_t$ is that there it is the value of the density $f\left(y_t|d\left(t-1\right),\hat{\theta}_{t-1}\right)$ with the substituted measured $y_t$ while here this density is used as a generator of $y_t$ from the same pdf[1].

## A multi-step prediction

If we want to predict several steps ahead, we must

- predict the pointer value to that point (power of $\alpha$),
- predict the output to that point from all components,
- combine the output predictions with the predicted weights from pointer prediction.

Remark

Do it in math and verify by program.

[1] At least, I hope it is like this.