# Doing Inference in a LR-HSMM

Kanru Hua

June 8, 2015

Derivation of LR-HSMM inference algorithms for the implementation of `liblrhsmm` is documented step-by-step here to the finest detail.

### Notations

| | |
|---|---|
| $o_t$ | observation at time $t$ (range for $t$: 1 to $T$) |
| $s_t = i$ | the state at time $t$ being $i$ (range for $i$: 1 to $i_{end}$) |
| $m_t = k$ | the emitting mixture at time $t$ being $k$ |
| $T_t$ | the event of a state starting from time $t$ |
| $b_i(\cdot)$ | observation probability function of state $i$ |
| $\delta_i(\cdot)$ | duration probability function of state $i$ |

### Forward Probability

$$\begin{aligned}
\alpha_t(i) &\triangleq \mathrm{P}(o_1, ..., o_t, s_t = i, T_t) \qquad (1)\\
&= \sum_{d=1}^{t-1} \mathrm{P}(o_1, ..., o_t, s_t = i, T_t, s_{t-d} = i-1, T_{t-d})\\
&= \sum_{d} \mathrm{P}(o_{t-d+1}, ..., o_t, s_t = i, T_t | \cancel{o_1, ..., o_{t-d}}, s_{t-d} = i-1, T_{t-d})\\
&\qquad \cdot \mathrm{P}(o_1, ..., o_{t-d}, s_{t-d} = i-1, T_{t-d})\\
&= \sum_{d} \frac{\mathrm{P}(o_{t-d+1}, ..., o_t, s_t = i, T_t, s_{t-d} = i-1, T_{t-d})}{\mathrm{P}(s_{t-d} = i-1, T_{t-d})} \cdot \mathrm{P}(o_1, ..., o_{t-d}, s_{t-d} = i-1, T_{t-d})\\
&= \sum_{d} \mathrm{P}(o_{t-d+1}, ..., o_t | s_t = i, T_t, s_{t-d} = i-1, T_{t-d}) \cdot \frac{\mathrm{P}(s_t = i, T_t, s_{t-d} = i-1, T_{t-d})}{\mathrm{P}(s_{t-d} = i-1, T_{t-d})}\\
&\qquad \cdot \mathrm{P}(o_1, ..., o_{t-d}, s_{t-d} = i-1, T_{t-d})\\
&= \sum_{d} \mathrm{P}(o_{t-d+1}, ..., o_t | s_t = i, T_t, s_{t-d} = i-1, T_{t-d}) \cdot \mathrm{P}(s_t = i, T_t | s_{t-d} = i-1, T_{t-d})\\
&\qquad \cdot \mathrm{P}(o_1, ..., o_{t-d}, s_{t-d} = i-1, T_{t-d})\\
&= b_i(o_t) \sum_{d} \left( \prod_{t'=t-d+1}^{t-1} b_{i-1}(o_{t'}) \right) \delta_{i-1}(d) \alpha_{t-d}(i-1)
\end{aligned}$$

$$\begin{aligned}
\alpha_1(1) &\triangleq \mathrm{P}(o_1, s_1 = 1, T_1) \qquad (2)\\
&= \mathrm{P}(o_1 | s_1 = 1, T_1) \cdot \underbrace{\mathrm{P}(s_1 = 1, T_1)}_{=1}\\
&= b_1(o_1)
\end{aligned}$$

### Backward Probability

$$
\begin{aligned}
\beta_t(i) &\triangleq \mathrm{P}(o_{t+1},..,o_T|s_t=i,T_t) \qquad (3)\\
&= \frac{\mathrm{P}(o_{t+1},...,o_T,s_t=i,T_t)}{\mathrm{P}(s_t=i,T_t)}\\
&= \sum_{d=1}^{T-t} \frac{\mathrm{P}(o_{t+1},...,o_T,s_t=i,T_t,s_{t+d}=i+1,T_{t+d})}{\mathrm{P}(s_t=i,T_t)}\\
&= \sum_d \frac{\mathrm{P}(o_{t+1},...,o_{t+d},s_t=i,T_t|\cancel{o_{t+d+1},...,o_T},s_{t+d}=i+1,T_{t+d})}{\mathrm{P}(s_t=i,T_t)}\\
&\qquad \cdot \mathrm{P}(o_{t+d+1},...,o_T,s_{t+d}=i+1,T_{t+d})\\
&= \sum_d \frac{\mathrm{P}(o_{t+1},...,o_{t+d},s_t=i,T_t,s_{t+d}=i+1,T_{t+d})}{\mathrm{P}(s_{t+d}=i+1,T_{t+d})\cdot \mathrm{P}(s_t=i,T_t)} \cdot \mathrm{P}(o_{t+d+1},...,o_T,s_{t+d}=i+1,T_{t+d})\\
&= \sum_d \mathrm{P}(o_{t+1},...,o_{t+d}|s_t=i,T_t,s_{t+d}=i+1,T_{t+d}) \cdot \frac{\mathrm{P}(s_t=i,T_t,s_{t+d}=i+1,T_{t+d})}{\mathrm{P}(s_t=i,T_t)}\\
&\qquad \cdot \frac{\mathrm{P}(o_{t+d+1},...,o_T,s_{t+d}=i+1,T_{t+d})}{\mathrm{P}(s_{t+d}=i+1,T_{t+d})}\\
&= \sum_d \mathrm{P}(o_{t+1},...,o_{t+d}|s_t=i,T_t,s_{t+d}=i+1,T_{t+d}) \cdot \mathrm{P}(s_{t+d}=i+1,T_{t+d}|s_t=i,T_t)\\
&\qquad \cdot \mathrm{P}(o_{t+d+1},...,o_T|s_{t+d}=i+1,T_{t+d})\\
&= \sum_d \left(\prod_{t'=t+1}^{t+d-1} b_i(o_{t'})\right) b_{i+1}(o_{t+d})\delta_i(d)\beta_{t+d}(i+1)
\end{aligned}
$$

$$
\begin{aligned}
\beta_t(i_{end}) &\triangleq \mathrm{P}(o_{t+1},...,o_T|s_t=i_{end},T_t), \quad \forall t\in[2,T] \qquad (4)\\
&= \mathrm{P}(o_{t+1},...,o_T|s_t=i_{end},T_t,T_{T+1})\\
&= \prod_{t'=t+1}^{T} b_{i_{end}}(o_{t'})
\end{aligned}
$$

### Total Probability

From Forward Probability

$$
\begin{aligned}
&\mathrm{P}(o_1,...,o_T) \qquad (5)\\
&= \sum_d \mathrm{P}(o_1,...,o_T,s_{T-d}=i_{end},T_{T-d})\\
&= \sum_d \mathrm{P}(o_1,...,o_{T-d}|s_{T-d}=i_{end},T_{T-d}) \cdot \mathrm{P}(o_{T-d+1},...,o_T|s_{T-d}=i_{end},T_{T-d})\\
&\qquad \cdot \mathrm{P}(s_{T-d}=i_{end},T_{T-d})\\
&= \sum_d \mathrm{P}(o_1,...,o_{T-d},s_{T-d}=i_{end},T_{T-d}) \cdot \mathrm{P}(o_{T-d+1},...,o_T|s_{T-d}=i_{end},T_{T-d})\\
&= \sum_d \alpha_{T-d}(i_{end}) \prod_{t'=T-d+1}^{T} b_{i_{end}}(o_{t'})
\end{aligned}
$$

From Forward and Backward Probability

$$\begin{aligned}
&\mathrm{P}(o_1, ..., o_T) \\
&= \sum_t \mathrm{P}(o_1, ..., o_T, s_t = i, T_t) \\
&= \sum_t \mathrm{P}(o_1, ..., o_T | s_t = i, T_t) \cdot \mathrm{P}(s_t = i, T_t) \\
&= \sum_t \mathrm{P}(o_1, ..., o_t | s_t = i, T_t) \cdot \mathrm{P}(o_{t+1}, ..., o_T | s_t = i, T_t) \cdot \mathrm{P}(s_t = i, T_t) \\
&= \sum_t \mathrm{P}(o_1, ..., o_t, s_t = i, T_t) \cdot \mathrm{P}(o_{t+1}, ..., o_T | s_t = i, T_t) \\
&= \sum_t \alpha_t(i) \beta_t(i)
\end{aligned} \tag{6}$$

**Consecutive States Occupancy Probability**

$$\begin{aligned}
\chi_{t,d}(i) &\triangleq \mathrm{P}(s_t = i, T_t, s_{t+d} = i+1, T_{t+d} | o_1, ..., o_T) \\
&= \mathrm{P}(o_1, ..., o_T, s_t = i, T_t, s_{t+d} = i+1, T_{t+d}) \cdot \frac{1}{\mathrm{P}(o_1, ..., o_T)} \\
&= \mathrm{P}(o_1, ..., o_T | s_t = i, T_t, s_{t+d} = i+1, T_{t+d}) \cdot \mathrm{P}(s_t = i, T_t, s_{t+d} = i+1, T_{t+d}) \cdot \frac{1}{\mathrm{P}(o_1, ..., o_T)} \\
&= \mathrm{P}(o_1, ..., o_t, s_t = i, T_t, s_{t+d} = i+1, T_{t+d}) \\
&\qquad \cdot \mathrm{P}(o_{t+1}, ..., o_{t+d} | s_t = i, T_t, s_{t+d} = i+1, T_{t+d}) \\
&\qquad \cdot \mathrm{P}(o_{t+d+1}, ..., o_T | s_t = i, T_t, s_{t+d} = i+1, T_{t+d}) \cdot \frac{1}{\mathrm{P}(o_1, ..., o_T)} \\
&= \mathrm{P}(o_1, ..., o_t | s_t = i, T_t, \cancel{s_{t+d} = i+1, T_{t+d}}) \cdot \mathrm{P}(s_t = i, T_t, s_{t+d} = i+1, T_{t+d}) \\
&\qquad \cdot \mathrm{P}(o_{t+1}, ..., o_{t+d} | s_t = i, T_t, s_{t+d} = i+1, T_{t+d}) \\
&\qquad \cdot \mathrm{P}(o_{t+d+1}, ..., o_T | \cancel{s_t = i, T_t}, s_{t+d} = i+1, T_{t+d}) \cdot \frac{1}{\mathrm{P}(o_1, ..., o_T)} \\
&= \mathrm{P}(o_1, ..., o_t, s_t = i, T_t) \cdot \mathrm{P}(s_{t+d} = i+1, T_{t+d} | s_t = i, T_t) \\
&\qquad \cdot \mathrm{P}(o_{t+1}, ..., o_{t+d} | s_t = i, T_t, s_{t+d} = i+1, T_{t+d}) \\
&\qquad \cdot \mathrm{P}(o_{t+d+1}, ..., o_T | s_{t+d} = i+1, T_{t+d}) \cdot \frac{1}{\mathrm{P}(o_1, ..., o_T)} \\
&= \frac{\alpha_t(i) \delta_i(d) \left( \prod_{t'=t+1}^{t+d-1} b_i(o_{t'}) \right) b_{i+1}(o_{t+d}) \beta_{t+d}(i+1)}{\mathrm{P}(o_1, ..., o_T)}
\end{aligned} \tag{7}$$

$$\begin{aligned}
\chi_{t,d}(i_{end}) &\triangleq \mathrm{P}(s_t = i_{end}, T_t, T_{t+d} | o_1, ..., o_T) \\
&= \mathrm{P}(o_1, ..., o_T | s_t = i_{end}, T_t, T_{t+d}) \cdot \mathrm{P}(s_t = i_{end}, T_t, T_{t+d}) \cdot \frac{1}{\mathrm{P}(o_1, ..., o_T)}
\end{aligned} \tag{8}$$

**Single State Occupancy Probability**

$$\begin{aligned}
\gamma_t(i) &\triangleq \mathrm{P}(s_t = i | o_1, ..., o_T) \\
&= \sum_{u=0}^{t-1} \sum_{v=1}^{T-t} \mathrm{P}(s_{t-u} = i, T_{t-u}, \cancel{s_t = i}, s_{t+v} = i+1, T_{t+v} | o_1, ..., o_T) \\
&= \sum_{u=0}^{t-1} \sum_{v=1}^{T-t} \mathrm{P}(s_{t-u} = i, T_{t-u}, s_{t+v} = i+1, T_{t+v} | o_1, ..., o_T) \\
&= \sum_{u=0}^{t-1} \sum_{v=1}^{T-t} \chi_{t-u,u+v}(i)
\end{aligned} \tag{9}$$

**Mixture Occupancy Probability (LR-CD-HSMM)**

$$\begin{aligned}
\gamma_t(i,k) &\triangleq \mathrm{P}(s_t = i, m_t = k | o_1, ..., o_T) \\
&= \mathrm{P}(m_t = k | o_1, ..., o_T, s_t = i) \cdot \mathrm{P}(s_t = i | o_1, ..., o_T) \\
&= \mathrm{P}(m_t = k | o_t, s_t = i) \cdot \mathrm{P}(s_t = i | o_1, ..., o_T) \\
&= \frac{\mathrm{P}(o_t | m_t = k, s_t = i) \cdot \mathrm{P}(m_t = k, s_t = i)}{\mathrm{P}(o_t, s_t = i)} \cdot \mathrm{P}(s_t = i | o_1, ..., o_T) \\
&= \frac{\mathrm{P}(o_t | m_t = k, s_t = i) \cdot \mathrm{P}(m_t = k | s_t = i)}{\mathrm{P}(o_t | s_t = i)} \cdot \mathrm{P}(s_t = i | o_1, ..., o_T) \\
&= \gamma_t(i) \frac{w_{ik} f_{ik}(o_t)}{b_i(o_t)}
\end{aligned} \tag{10}$$