1 Filtering

1.1 Noisy observation of a markov chain (hidden Markov chain)

Assume X_n is a time-homogeneous Markov chain on S with a transition prob. function $p(c,x)$. Consider a sequence of noisy observations

$$Y_n = \gamma(X_n, W_n), n \geq 0,$$

where $W_n, n \geq 0$, are a sequence of i.i.d. Set $l(u,v) = \mathbf{P}(\gamma(u, W_0) = v)$. Then $Z_n = (X_n, Y_n), n \geq 0$, is a homogeneous Markov chain. We find its transition probabilities

$$\begin{aligned}
& \mathbf{P}(X_{n+1} = u, Y_{n+1} = v | X_n = r, Y_n = s) \\
= \;& \mathbf{P}(X_{n+1} = u, \gamma(X_{n+1}, W_{n+1}) = v | X_n = r, Y_n = s) \\
= \;& \mathbf{P}(X_{n+1} = u, \gamma(u, W_{n+1}) = v | X_n = r, Y_n = s) \\
= \;& \mathbf{P}(\gamma(u, W_{n+1}) = v)\, \mathbf{P}(X_{n+1} = u | X_n = r, Y_n = s) \\
= \;& l(u,v)\mathbf{P}(X_{n+1} = u | X_n = r, \gamma(X_n, W_n) = s) = q(u,v)\mathbf{P}(X_{n+1} = u | X_n = r, \gamma(r, W_n) = s) \\
= \;& l(u,v)p(r,u).
\end{aligned}$$

1.2 Filtering of a Markov chain

Suppose $Z_n = (X_n, Y_n), n \geq 0$, is a homogeneous Markov chain on $S = A \times B$ with the transition probability $p((r,s),(u,v)), (r,s),(u,v) \in S$. We will denote $q^{r,u}(s,v) = p((r,s),(u,v))$ and assume it is known.

Filtering problem: Given $Y_0, \ldots, Y_n$, find the best estimate of $f(X_n)$, where $n = 0, 1, \ldots$.; or given $Y_{[0,n]} = (Y_0, \ldots, Y_n) = b_{[0,n]} = (b_0, \ldots, b_n)$, find the best estimate of $f(X_n)$.

The answer is $\mathbf{E}(f(X_n)|Y_{[0.n]})$ or $\mathbf{E}(f(X_n)|Y_{[0.n]} = b_{[0,n]})$.

All these quantities are determined by the collection of conditional distributions

$$\pi_n^r(b_{[0,n]}) = \mathbf{P}\left(X_n = r | Y_{[0,n]} = b_{[0,n]}\right), r \in A, b_{[0,n]} = (b_0, \ldots, b_n) \in B^n, n \geq 0.$$

Since

$$\pi_n^r(b_{[0,n]}) = \mathbf{P}\left(X_n = r | Y_{[0,n]} = b_{[0,n]}\right) = \frac{\mathbf{P}\left(X_n = r, Y_{[0,n]} = b_{[0,n]}\right)}{\mathbf{P}\left(Y_{[0,n]} = b_{[0,n]}\right)},$$

(convention $c/0 = 0$ holds) and $\mathbf{P}\left(Y_{[0,n]} = b_{[0,n]}\right) = \sum_{r \in A} \mathbf{P}\left(X_n = r, Y_{[0,n]} = b_{[0,n]}\right)$, we start with computation of the joint probabilities

$$\phi_n^r\left(b_{[0,n]}\right) = \mathbf{P}\left(X_n = r, Y_{[0,n]} = b_{[0,n]}\right), r \in A, b_{[0,n]} \in B^{n+1}, n \geq 0.$$

For $n = 0$, we have $\phi_n^r\left(b_{[0,0]}\right) = \phi_0^r(b) = \mathbf{P}(X_0 = r, Y_0 = b), r \in A, b \in B$, and we assume that it is known (given). We compute $\phi_n^r\left(b_{[0,n]}\right)$ step by step.

Lemma 1 *For any $a \in A, n \geq 0$ and $b_0, b_1, \ldots,$*

$$\phi_{n+1}^{a}\left(b_{[0,n+1]}\right) = \sum_{r\in A} q^{r,a}(b_n, b_{n+1})\phi_n^r\left(b_{[0,n]}\right)$$

Comment. Lemma states that the joint probabilities $\phi_n^r\left(b_{[0,n]}\right)$ can be computed step by step starting with $\phi_0^r(b)$: $\phi_{n+1}^r\left(b_{[0,n+1]}\right)$ computed using $q^{r,u}(s,v)$ (which is known) and the joint probabilities $\phi_n^r\left(b_{[0,n]}\right)$ found before.
Proof. By the total probability formula and Markov property,

$$\begin{aligned}
\phi_{n+1}^{a}\left(b_{[0,n+1]}\right) &= \mathbf{P}\left(X_{n+1}=a, Y_{[0,n+1]}=b_{[0,n+1]}\right)\\
&= \mathbf{P}\left(X_{n+1}=a, Y_{n+1}=b_{n+1}, Y_{[0,n]}=b_{[0,n]}\right)\\
&= \sum_{r\in A}\mathbf{P}\left(X_{n+1}=a, Y_{n+1}=b_{n+1}, X_n=r, Y_{[0,n]}=b_{[0,n]}\right)\\
&= \sum_{r\in A}\mathbf{P}\left(X_{n+1}=a, Y_{n+1}=b_{n+1}|X_n=r, Y_{[0,n]}=b_{[0,n]}\right)\mathbf{P}\left(X_n=r, Y_{[0,n]}=b_{[0,n]}\right)\\
&= \sum_{r\in A} q^{r,a}(b_n, b_{n+1})\phi_n^r\left(b_{[0,n]}\right).
\end{aligned}$$

■

Note, that by the total probability formula,

$$\mathbf{P}\left(Y_{[0,n]}=b_{[0,n]}\right) = \sum_{u\in A}\mathbf{P}\left(X_n=u, Y_{[0,n]}=b_{[0,n]}\right) = \sum_{u\in A}\phi_n^u\left(b_{[0,n]}\right).$$

So, we are ready to find the conditional probabilities

$$\begin{aligned}
\pi_n^a(b_{[0,n]}) &= \mathbf{P}\left(X_n=a|Y_{[0,n]}=b_{[0,n]}\right) = \frac{\mathbf{P}\left(X_n=a, Y_{[0,n]}=b_{[0,n]}\right)}{\mathbf{P}\left(Y_{[0,n]}=b_{[0,n]}\right)}\\
&= \frac{\phi_n^a\left(b_{[0,n]}\right)}{\sum_{u\in A}\phi_n^u\left(b_{[0,n]}\right)}, n\geq 0.
\end{aligned}$$

Already found? We prefer step by step formulas assuming again that $\pi_0^a(b) = \mathbf{P}\left(X_0=a|Y_0=b\right), a\in A, b\in B$, are known.

Theorem 2 *For any $a \in A, n \geq 0$ and $b_0, b_1, \ldots,$*

$$\pi_{n+1}^a(b_{[0,n+1]}) = \frac{\sum_{r\in A} q^{r,a}(b_n,b_{n+1})\pi_n^r(b_{[0,n]})}{\sum_{u\in A}\left(\sum_{r\in A} q^{r,u}(b_n,b_{n+1})\pi_n^r(b_{[0,n]})\right)};$$

Proof. By Lemma 1,

$$\begin{aligned}
\pi_{n+1}^a(b_{[0,n+1]}) &= \frac{\phi_{n+1}^a\left(b_{[0,n+1]}\right)}{\sum_{u\in A}\phi_{n+1}^u\left(b_{[0,n+1]}\right)} = \frac{\sum_{r\in A} q^{r,a}(b_n,b_{n+1})\phi_n^r\left(b_{[0,n]}\right)}{\sum_{u\in A}\sum_{r\in A} q^{r,u}(b_n,b_{n+1})\phi_n^r\left(b_{[0,n]}\right)}\\
&= \frac{\sum_{r\in A} q^{r,a}(b_n,b_{n+1})\frac{\phi_n^r\left(b_{[0,n]}\right)}{\mathbf{P}\left(Y_n=b_{[0,n]}\right)}}{\sum_{u\in A}\sum_{r\in A} q^{r,u}(b_n,b_{n+1})\frac{\phi_n^r\left(b_{[0,n]}\right)}{\mathbf{P}\left(Y_n=b_{[0,n]}\right)}} = \frac{\sum_{r\in A} q^{r,a}(b_n,b_{n+1})\pi_n^r(b_{[0,n]})}{\sum_{u\in A}\left(\sum_{r\in A} q^{r,u}(b_n,b_{n+1})\pi_n^r(b_{[0,n]})\right)}.
\end{aligned}$$

■

The meaning of the quotient: in the formula for $\pi^a_{n+1}(b_{[0,n+1]}) = \mathbf{P}\left(X_{n+1} = a | Y_{[0,n+1]} = b_{[0,n+1]}\right)$,

$$\begin{aligned}
\text{numerator} &= \sum_{r\in A} q^{r,a}(b_n, b_{n+1})\phi^r_n\left(b_{[0,n]}\right) = \mathbf{P}\left(X_{n+1} = a, Y_{n+1} = b_{n+1} | Y_{[0,n]} = b_{[0,n]}\right), \\
\text{denominator} &= \xi_{n+1}\left(b_{[0,n+1]}\right) = \sum_{u\in A}\left(\sum_{r\in A} q^{r,u}(b_n, b_{n+1})\pi^r_n(b_{[0,n]})\right) = \mathbf{P}\left(Y_{n+1} = b_{n+1} | Y_{[0,n]} = b_{[0,n]}\right).
\end{aligned}$$

Remark 3 *1) The function $\phi^a_n = \phi^a_n(Y_{[0,n]}), a \in A$, is called unnormalized filtering distribution (or density) function (UFDF), and $\pi^a_n = \pi^a_n(Y_{[0,n]}), a \in A$, is called a filtering distribution (or desnsity) function (FDF); the relation berween them:*

$$\pi^a_n = \frac{\phi^a_n}{\sum_{r\in A}\phi^r_n}, a \in A; \tag{1}$$

2) In Lemma 1 and Theorem 2 we found that for every $a \in A, n \geq 0, b_0, \ldots \in B$,

$$\phi^a_{n+1} = \sum_r q^{r,a}(Y_n, Y_{n+1})\phi^r_n, \tag{2}$$

and

$$\pi^a_{n+1} = \frac{\sum_{r\in A} q^{r,a}(Y_n, Y_{n+1})\pi^r_n}{\sum_{u\in A}\sum_{r\in A} q^{r,u}(Y_n, Y_{n+1})\pi^r_n}; \tag{3}$$

Formula (2) combined with 1) is preferable to (3): division errors do not accumulate);

3) The best mean square estimate of $f(X_n)$ is

$$\hat{f(X_n)} = \mathbf{E}\left[f(X_n)|Y_{[0,n]}\right] = \sum_{a\in A} f(a)\pi^a_n = \frac{\sum_{a\in A} f(a)\phi^a_n}{\sum_{a\in A}\phi^r_n};$$

4) If $A = \{a_1, \ldots, a_d\}$ is finite, then (2) can be rewritten in matrix form. Define

$$\begin{aligned}
q^{ij}_n &= q^{a_i,a_j}(Y_n, Y_{n+1}), 1 \leq i, j \leq d, \\
\phi^i_n &= \phi^{a_i}_n, 1 \leq i \leq d, \\
q_n &= \left(q^{ij}_n\right)_{1\leq i,j\leq n}, \phi_n = \left(\phi^i_n\right)_{1\leq i\leq d}.
\end{aligned}$$

Using these notations (2) becomes

$$\begin{aligned}
\phi^j_{n+1} &= \sum_{i=1}^d \phi^i_n q^{ij}_n, j = 1, \ldots, d, \\
&\quad or \\
\phi_{n+1} &= \phi_n q_n, n \geq 0.
\end{aligned}$$

Example 4 *(Hidden Markov Chain) Assume X_n is a homogeneous Markov chain with transition probabilities function $p(r,u)$. Let*

$$Y_n = \gamma(X_n, W_n), n \geq 0,$$

where $W_0, W_1, \ldots$ are i.i.d. indepensdent of $X_0, X_1, \ldots$. Denote $l(u,v) = \mathbf{P}(\gamma(u, W_1) = v)$. Then we found that

$$q^{r,u}(s,v) = p(r,u)l(u,v)$$

does not depenfd on s. If $A = \{a_1, \ldots, a_d\}$ is finite, then , denoting $p_{ij} = p(a_i, a_j), l_j(v) = l(a_j, v)$, and following Remark 3 4) we have

$$q_n^{ij} = q^{a_i, a_j}(Y_n, Y_{n+1}) = p_{ij} l_j(Y_{n+1}), 1 \leq i, j \leq d,$$

and (2) becomes

$$\phi_{n+1}^j = l_j(Y_{n+1}) \sum_{i=1}^d \phi_n^i p_{ij}, j = 1, \ldots, d.$$

1.3 Space continuous observation in HMM (hidden Markov chain model)

Signal: $X_n, n \geq 0$, is a homogeneous Markov chain on A with transition probabilities $p(r,u)$;

Observation is defined by

$$Y_n = h(X_n) + W_n, n \geq 0,$$

where $h(x)$ is a real-valued function on A and W_n are i.i.d. continuous r.v. with pdf $l(v), v \in \mathbf{R}$. For example, if $W_n \sim N(0, \sigma^2)$, then

$$l(v) = \frac{1}{\sqrt{2\pi}\sigma} e^{-v^2/2\sigma^2}, v \in \mathbf{R}.$$

Problem: compute filtering density function

$$\pi_n^a = \mathbf{P}\left(X_n = a | Y_{[0,n]}\right), a \in A.$$

According to Lemma 14 in Appendix $\pi_n^a = \pi_n^a(b_{[0,n]})|_{b_{[0,n]} = Y_{[0,n]}}$, where

$$\pi_n^a\left(b_{[0,n]}\right) = \frac{\phi_n^a(b_{[0,n]})}{\sum_{v \in A} \phi_n^v(b_{[0,n]})},$$

and

$$\phi_n^a(b_{[0,n]}) = \frac{\partial^{n+1}}{\partial b_0 \ldots \partial b_n} \mathbf{P}\left(X_n = a, Y_{[0,n]} \leq b_{[0,n]}\right). \tag{4}$$

Obviously, $\phi_0^a(b_0) = \mathbf{P}(X_0 = a)\, l(b_0 - h(a)), a \in A, b_0 \in \mathbf{R}$. The following step by step formula holds for UFDF $\phi_n^a(b_{[0,n]})$

Proposition 5 *For any* $a \in A, n \geq 0, b_0, \ldots \in \mathbf{R}$,

$$\phi^a_{n+1}(b_{[0,n+1]}) = l\left(b_{n+1} - h(a)\right) \sum_{r \in A} p(r,a)\phi^r_n(b_{[0,n]}). \tag{5}$$

Proof. Since (4) holds, we consider

$$\mathbf{P}\left(X_{n+1} = a, Y_{[0,n+1]} \leq b_{[0,n+1]}\right) = \sum_{r \in A} \mathbf{P}\left(X_{n+1} = a, Y_{n+1} \leq b_{n+1}, X_n = r, Y_{[0,n]} \leq b_{[0,n]}\right).$$

We have, denoting $H(X_{[0,n]}) = (h(X_0), \ldots, h(X_n))$,

$$\begin{aligned} & \mathbf{P}\left(X_{n+1} = a, Y_{n+1} = b_{n+1}, X_n = r, Y_{[0,n]} \leq b_{[0,n]}\right) \\ = \; & \mathbf{P}\left(X_{n+1} = a, W_{n+1} \leq b_{n+1} - h(a), X_n = r, H\left(X_{[0,n]}\right) + W_{[0,n]} \leq b_{[0,n]}\right) \\ = \; & \mathbf{P}\left(W_{n+1} \leq b_{n+1} - h(a)\right) \mathbf{P}\left(X_{n+1} = a, X_n = r, H\left(X_{[0,n]}\right) + W_{[0,n]} \leq b_{[0,n]}\right), \end{aligned}$$

and (W and X are independent)

$$\begin{aligned} & \mathbf{P}\left(X_{n+1} = a, X_n = r, H\left(X_{[0,n]}\right) + W_{[0,n]} \leq b_{[0,n]}\right) \\ = \; & \mathbf{P}\left(X_{n+1} = a, X_n = r, H\left(X_{[0,n]}\right) + c \leq b_{[0,n]}\right)|_{c = W_{[0,n]}}. \end{aligned}$$

Finally, by condtioning

$$\begin{aligned} & \mathbf{P}\left(X_{n+1} = a, X_n = r, H\left(X_{[0,n]}\right) + c \leq b_{[0,n]}\right) \\ = \; & \mathbf{P}\left(X_{n+1} = a | X_n = r, H\left(X_{[0,n]}\right) + c \leq b_{[0,n]}\right) \mathbf{P}\left(X_n = r, H\left(X_{[0,n]}\right) + c \leq b_{[0,n]}\right) \\ = \; & \mathbf{P}\left(X_{n+1} = a | X_n = r\right) \mathbf{P}\left(X_n = r, H\left(X_{[0,n]}\right) + c \leq b_{[0,n]}\right) \\ = \; & p(r,a)\mathbf{P}\left(X_n = r, H\left(X_{[0,n]}\right) + c \leq b_{[0,n]}\right). \end{aligned}$$

So,

$$\begin{aligned} & \mathbf{P}\left(X_{n+1} = a, X_n = r, H\left(X_{[0,n]}\right) + W_{[0,n]} \leq b_{[0,n]}\right) \\ = \; & p(r,a)\mathbf{P}\left(X_n = r, H\left(X_{[0,n]}\right) + W_{[0,n]} \leq b_{[0,n]}\right) \\ = \; & p(r,a)\mathbf{P}\left(X_n = r, Y_{[0,n]} \leq b_{[0,n]}\right) \end{aligned}$$

and the statement follows by taking the mixed derivative $\partial^{n+2}/\partial b_0 \ldots \partial b_{n+1}$ of both sides in

$$\mathbf{P}\left(X_{n+1} = a, Y_{[0,n+1]} \leq b_{[0,n+1]}\right) = \sum_{r \in A} \mathbf{P}\left(W_{n+1} \leq b_{n+1} - h(a)\right) p(r,a)\mathbf{P}\left(X_n = r, Y_{[0,n]} \leq b_{[0,n]}\right).$$

$\blacksquare$

Remark 6 *1) Denotning* $\phi_n^a = \phi_n^a(Y_{[0,n]}), a \in A$, *(UFDF), and* $\pi_n^a = \pi_n^a(Y_{[0,n]}), a \in A$, *(FDF) like in the space discrete case*

$$\pi_n^a = \frac{\phi_n^a}{\sum_{u \in A} \phi_n^u}, a \in A;$$

and we can rewrite (5) as

$$\phi_{n+1}^a = l(Y_{n+1} - h(a)) \sum_{r \in A} p(r,a)\phi_n^r, a \in A.$$

Also,

$$\pi_{n+1}^a = \frac{l(Y_{n+1} - h(a)) \sum_{r \in A} p(r,a)\pi_n^r}{\sum_{u \in A} \left(l(Y_{n+1} - h(u)) \sum_{r \in A} p(r,u)\pi_n^r \right)}, u \in A.$$

Again the formula for UFDF is preferrable to FDF.

2) If $A = \{a_1, \ldots, a_d\}$ *is finite, then they can be rewritten in matrix form. Define* $p_{ij} = p(a_i, a_j), 1 \le i, j \le d$,

$$\begin{aligned} q_n^{ij} &= l(Y_{n+1} - h(a_j))p_{ij}, 1 \le i, j \le d, \\ \phi_n^i &= \phi_n^{a_i}, 1 \le i \le d, \\ q_n &= \left(q_n^{ij}\right)_{1 \le i,j \le n}, \phi_n = \left(\phi_n^i\right)_{1 \le i \le d}. \end{aligned}$$

The equation for UFDF becomes

$$\begin{aligned} \phi_{n+1}^j &= l(Y_{n+1} - h(a_j)) \sum_{i=1}^d p_{ij}\phi_n^i, j = 1, \ldots, d, \\ &\quad \textit{or} \\ \phi_{n+1} &= \phi_n q_n, n \ge 0. \end{aligned}$$

2 Smoothing and Prediction for HMM

Consider a hidden Markov chain model (HMM).

Signal is a homogeneous Markov chain on $A = \{a_1, a_2, \ldots\}$, with the transition probabilities $p(r, u)$;

Observation $Y_n, n \geq 0$, is given by

$$Y_n = \gamma(X_n, W_n), \tag{6}$$

where $W_0, W_1 \ldots$ are discrete i.i.d. independent of the signal $X_n, n \geq 0$;

Also, we will consider a continuous space noise model where the observation

$$Y_n = h(X_n) + W_n, n \geq 0, \tag{7}$$

where $W_0, W_1 \ldots$ are continuous i.i.d. with a given pdf $l(v)$ independent of the signal $X_n, n \geq 0$;

We introduce the function

$$l(u, v) = \begin{cases} \mathbf{P}\left(\gamma(u, W_1) = v\right), & \text{in the case (6)}, \\ \quad l(v - h(u)), & \text{in the case (7)}. \end{cases}$$

(we assume it is known).

The joint pdf of X_n and $Y_{[0,n]} = (Y_0, \ldots, Y_n)$ is

$$\phi_n^a(b_{[0,n]}) = \begin{cases} \mathbf{P}(X_n = a, Y_{[0,n]} = b_{[0,n]}), & \text{in the case (6)}, \\ \frac{\partial^{n+1}\mathbf{P}(X_n = a, Y_{[0,n]} \leq b_{[0,n]})}{\partial b_0 \ldots \partial b_n}, & \text{in the case (7)}, \end{cases}$$

where $a \in A, b_{[0,n]} = (b_0, \ldots, b_n) \in B^{n+1}$ (B is the set where $Y_n$takes its values).

Remark 7 *1) Obviously,* $\phi_0^a(b_0) = l(a, b_0)\mathbf{P}(X_0 = a)$;

2) Knowledge of $\phi_n^a = \phi_n^a\left(Y_{[0,n]}\right), a \in A,$ *(called unnormalized filtering density function UFDF) allows to find the conditional distributions*

$$\pi_n^a = \pi_n^a\left(Y_{[0,n]}\right) = \mathbf{P}\left(X_n = a | Y_{[0,n]}\right) = \frac{\phi_n^a}{\sum_{u \in A} \phi_n^u}, a \in A,$$

(called filtering density function, FDF).

We want to address the problem of prediction of the state X_T given the observation $Y_{[0,n]}, n \leq T$, and smoothing, estimation of the state $X_n, n \leq T$, given the observation $Y_{[0,T]}$.

We start with the smoothing problem. Since we want to find the conditional distributions

$$\mathbf{P}\left(X_n = a | Y_{[0,T]}\right), a \in A,$$

we need to consider the joint density function of X_n and $Y_{[0,T]}$:

$$\phi_{n,T}^a\left(b_{[0,T]}\right) = \begin{cases} \mathbf{P}\left(Y_{[0,T]} = b_{[0,T]}, X_n = a\right) \text{ in discrete case}, \\ \\ \partial_{b_{[0,T]}}^{T+1}\mathbf{P}\left(Y_{[0,T]} \leq b_{[0,T]}, X_n = a\right) \text{ in continuous case}. \end{cases}$$

The stochastic process $\boldsymbol{\phi}^a_{n,T} = \phi^a_{n,T}\left(Y_{[0,T]}\right), a \in A$, will be referred to as the *unnormalized smoothing distribution* (resp. *density*) function of the state process X_n given observation $Y_{[0,T]} = (Y_0, Y_1, ..., Y_T)$ in discrete (resp. continuous) case. In both cases unless there is a risk of confusion, we will use the abbreviation USDF.

Proposition 8 *In the discrete case, for all $a \in A, b_0, \ldots, b_T \in B, n \in [0, T]$,*

$$\phi^a_{n,T}\left(b_{[0,T]}\right) = \phi^a_n\left(b_{[0,n]}\right)\mu^a_n\left(b_{[n+1,T]}\right), \tag{8}$$

where $\mu^a_n\left(b_{[n+1,T]}\right) = \mathbf{P}\left(Y_{[n+1,T]} = b_{[n+1,T]}|X_n = a\right)$, if $n < T$, and $\mu^a_n\left(b_{[n+1,T]}\right) = \mu^a_T = 1$, if $n = T$.

The same relation holds in the continuous case as well.

***Note** that for every $n \geq 0$*

$$\sum_{a\in A}\phi^a_n\left(b_{[0,n]}\right)\mu^a_n\left(b_{[n+1,T]}\right) = \mathbf{P}\left(Y_{[0,T]} = b_{[0,T]}\right).$$

Proof. Indeed,since W and X are independent, by Lemma 16 (see Appendix)

$$\begin{aligned}
\phi^a_{n,T}\left(b_{[0,T]}\right) &= \mathbf{P}\left(Y_{[0,T]} = b_{[0,T]}, X_n = a\right) \\
&= \mathbf{P}\left(Y_{[0,n]} = b_{[0,n]}, X_n = a, Y_{[n+1,T]} = b_{[n+1,T]}\right) \\
&= \mathbf{P}(Y_{[n+1,T]} = b_{[n+1,T]}|Y_{[0,n]} = b_{[0,n]}, X_n = a)\phi^a_n\left(b_{[0,n]}\right) \\
&= \phi^a_n\left(b_{[0,n]}\right)\mu^a_n\left(b_{[n+1,T]}\right).
\end{aligned}$$

Obviously,

$$\begin{aligned}
\sum_{a\in A}\phi^a_n\left(b_{[0,n]}\right)\mu^a_n\left(b_{[n+1,T]}\right) &= \sum_{a\in A}\phi^a_{n,T}\left(b_{[0,T]}\right) = \sum_{a\in A}\mathbf{P}\left(X_n = a, Y_{[0,T]} = b_{[0,T]}\right) \\
&= \mathbf{P}\left(Y_{[0,T]} = b_{[0,T]}\right)
\end{aligned}$$

■

2.1 Backward Baum-Welch equation

We know already that $\phi^a_n\left(b_{[0,n]}\right)$ satisfies the forward Baum-Welch equation. We will show that $\mu^a_n\left(b_{[n+1,T]}\right)$ can be found by solving a backward Baum-Welcj equation.

FORWARD BAUM-WELCH equation: for all $a \in A, n \geq 0, b_0, \ldots \in B$,

$$\phi^a_{n+1}\left(b_{[0,n+1]}\right) = l(a, b_{n+1})\sum_{r\in A}p(r,a)\phi^r_n(b_{[0,n]}), \tag{9}$$

and for $\phi^a_n = \phi^a_n\left(Y_{[0,n]}\right)$,

$$\phi^a_{n+1} = l(a, Y_{n+1})\sum_{r\in A}p(r,a)\phi^r_n. \tag{10}$$

Recall, $\phi_n^a\left(b_{[0,n]}\right), a \in A, b_{[0,n]} \in B^{n+1}$ is the joint pdf of X_n and $Y_{[0,n]}$.

BACKWARD BAUM-WELCH equation. It is an equation for

$$\mu_n^a(b_{[n+1,T]}) = \begin{cases} \mathbf{P}\left(Y_{[n+1,T]} = b_{[n+1,T]} | X_n = a\right), & \text{in the case (6)} \\ \partial^{T-n}\mathbf{P}\left(Y_{[n+1,T]} \leq b_{[n+1,T]} | X_n = a\right)/\partial b_{n+1} \ldots \partial b_T & \text{in the case (7)} \end{cases}$$

The following statement regarding HMM (with space discrete or space continuous noise) holds.

Lemma 9 *For all $a \in A$ and $b_T, b_{T-1}, \ldots, \in B, T-1 > n \geq 0$,*

$$\mu_n^a(b_{[n+1,T]}) = \sum_{r \in A} l(r, b_{n+1}) p(a,r) \mu_{n+1}^r(b_{[n+2,T]}),$$

and

$$\mu_{T-1}^a(b_T) = \sum_{r \in A} l(r, b_T) p(a,r) = \sum_{r \in A} l(r, b_T) p(a,r) \mu_T^r,$$

(recall we defined $\mu_T^r = 1$ for all r)

It is called a backward Baum-Welch equation.

Proof. We prove it for the space discrete case. We have $\mu_n^a(b_{[n+1,T]}) =$

$$\begin{aligned} &= \mathbf{P}\left(Y_{[n+1,T]} = b_{[n+1,T]} | X_n = a\right) \\ &= \sum_{r \in A} \mathbf{P}\left(Y_{[n+2,T]} = b_{[n+2,T]}, Y_{n+1} = b_{n+1}, X_{n+1} = r | X_n = a\right). \qquad (11) \end{aligned}$$

Since $Y_{[n+2,T]} = \Gamma(X_{[n+2,T]}, W_{[n+2.T]}) = (\gamma(X_{n+2}, W_{n+2}), \ldots, \gamma(X_T, W_T))$ (W and X are independent), we get

$$\begin{aligned} &\mathbf{P}\left(Y_{[n+2,T]} = b_{[n+2,T]}, Y_{n+1} = b_{n+1}, X_{n+1} = r | X_n = a\right) \\ =\ &\mathbf{P}\left(Y_{[n+2.T]} = b_{[n+2,T]}, \gamma(r, W_{n+1})) = b_{n+1}, X_{n+1} = r | X_n = a\right) \qquad (12) \\ =\ &\mathbf{P}\left(\gamma(r, W_{n+1})) = b_{n+1}\right) \mathbf{P}\left(Y_{[n+2.T]}) = b_{[n+2,T]}, X_{n+1} = r | X_n = a\right), \end{aligned}$$

and by Lemma 16

$$\begin{aligned} &\mathbf{P}\left(Y_{[n+2,T]} = b_{[n+2,T]}, X_{n+1} = r | X_n = a\right) \\ =\ &\mathbf{P}\left(Y_{[n+2,T]} = b_{[n+2,T]} | X_{n+1} = r, X_n = a\right) \mathbf{P}\left(X_{n+1} = r | X_n = a\right) \quad (13) \\ =\ &\mathbf{P}\left(Y_{[n+2,T]} = b_{[n+2,T]} | X_{n+1} = r\right) p(a,r) = \mu_{n+1}^r(b_{[n+2,T]}) p(a,r). \end{aligned}$$

We get the equation by summarizing (11)-(13).

Obviously, Proposition 8 implies immediately the following ■

Corollary 10 *Forward and backward Baum-Welch equations are in duality, i.e. for every n,*

$$\left(\boldsymbol{\mu}_n\left(b_{[n+1,T]}\right), \boldsymbol{\phi}_n\left(b_{[0,n]}\right)\right) := \sum_{a \in A} \mu_n^a\left(b_{[n+1,T]}\right) \phi_n^a\left(b_{[0,n]}\right) = \mathbf{P}\left(Y_{[0,T]} = b_{[0,T]}\right).$$

2.2 Prediction and Smoothing

Consider random function $\phi^a_{n,T} = \phi^a_{n,T}\left(Y_{[0,T]}\right).$

Proposition 11 *For every $a \in A$, the conditional probability distribution of the state X_n given the complete observation sequence*

$$\pi^a_{n,T} = \mathbf{P}\left(X_n = a | Y_{[0,T]}\right) = \frac{\phi^a_{n,T}}{\sum_{r\in A}\phi^r_T(Y_{[0,T]})}.$$

Note $\sum_{r\in A}\phi^r_T(b_{[0,T]}) = \sum_{r\in A}\mathbf{P}\left(X_n = r, Y_{[0,T]} = b_{[0,T]}\right) = \mathbf{P}\left(Y_{[0,T]} = b_{[0,T]}\right).$

Proof. Indeed,

$$\mathbf{P}\left(X_n = a | Y_{[0,T]} = b_{[0,T]}\right) =$$

$$\Pr\left(X_n = a, Y_{[0,T]} = b_{[0,T]}\right) / \mathbf{P}\left(Y_{[0,T]} = b_{[0,T]}\right).$$

■

Now, we address the prediction problem.

For $n < T$, write

$$\phi^a_{T,n}\left(b_{[0,n]}\right) = \begin{cases} \mathbf{P}\left(Y_{[0,n]} = b_{[0,n]}, X_T = a\right) & \text{in discrete case,} \\ \partial^{n+1}_{b_{[0,n]}}\mathbf{P}\left(Y_{[0,n]} \le b_{[0,n]}, X_T = a\right) & \text{in continuous case.} \end{cases}$$

The stochastic process $\boldsymbol{\phi}^a_{T,n}, a \in A$, will be referred to as the *unnormalized prediction distribution* (resp. *density*) function of the state process X_T given observation $Y_{[0,n]} = (Y_0, ..., Y_n)$ in discrete (resp. continuous) case. In both cases unless there is a risk of confusion, we will use the abbreviation UPDF.

Proposition 12 *For $n < T$,*

$$\phi^a_{T,n}\left(b_{[0,n]}\right) = \sum_{r\in A}\phi^r_n\left(b_{[0,n]}\right)\mathbf{P}\left(X_T = a | X_n = r\right), \tag{14}$$

and we find $\pi^a_{T,n}(b_{[0,n]}) = \mathbf{P}\left(X_T = a | Y_{[0,n]} = b_{[0,n]}\right)$ by

$$\pi^a_{T,n}(b_{[0,n]}) = \frac{\phi^a_{T,n}\left(b_{[0,n]}\right)}{\sum_{r\in}\phi^r_n(b_{[0,n]})} = \frac{\sum_{r\in A}\phi^r_n\left(b_{[0,n]}\right)\mathbf{P}\left(X_T = a | X_n = r\right)}{\sum_{r\in A}\phi^r_n\left(b_{[0,n]}\right)}.$$

Remark 13 *In matrix notation, equation* (14) *is given by $\boldsymbol{\phi}_{T,n} = \boldsymbol{\phi}_n\left(P^*\right)^{T-n}$ where P is the transition probability matrix of the Markov chain (X_n).*

Proof. Proof. It is readily checked that

$$\begin{array}{c} \phi^a_{T,n}\left(b_{[0,n]}\right) = \mathbf{P}\left(Y_{[0,n]} = b_{[0,n]}, X_T = a\right) \\ = \sum_{r\in A}\mathbf{P}\left(Y_{[0,n]} = b_{[0,n]}, X_n = r, X_T = a\right) \\ = \sum_{r\in A}\mathbf{P}\left(X_T = a | Y_{[0,n]} = b_{[0,n]}, X_n = r\right)\mathbf{P}\left(Y_{[0,n]} = b_{[0,n]}, X_n = r\right), \end{array} \tag{15}$$

and the statement follows by Lemma 16. ■

2.3 Problems.

1. 1. Consider an HMM representation of coin tossing experiment. Assume a three-state model (corresponding to three different coins) with probabilities

	State 1	State 2	State 3
$P(H)$	0.5	0.75	0.25
$P(T)$	0.5	0.25	0.75

   and with all state-transition probabilities equal to 1/3. Assume initial state probabilities of 1/3.

   You observe the sequence

   $Y^{10} = (HHHTHTHTTH)$.

   Using Baum Eq.'s, COMPUTE:

   $P(H$ in the fifth toss$|\ Y^8)$

   $P(H$ in the forth toss $|Y^{10})$

   $P(T$ in the $n-$th toss $|Y^n)$ for $n = 4, 5, 6$.

2. Consider an HMM. Do not assume memoryless channel. Prove that Y_n is independent of (Y_{n-1}, X_{n-1}) given X_n (i.e. $P(Y_n, Y_{n-1}, X_{n-1}|X_n) = P(Y_n|X_n)P(Y_{n-1}, X_{n-1}|X_n))$ if and only if $P(Y_n|X_n, Y_{n-1}, X_{n-1}) = P(Y_n|X_n)$.

3 Appendix

3.1 Conditioning

3.1.1 Conditional density

Assume V is a discrete r.v. with values in A and $U = (U_1, \ldots, U_d)$ is a $\mathbf{R}^d$-valued continuous r.v. with a joint pdf $g(v, u)$ (which means that for any bounded $f(v, u)$,

$$\mathbf{E}f\left(V, U\right) = \sum_{v \in A} \int_{\mathbf{R}^d} f(v, u)g(v, u)\ du).$$

Note that

$$g(v, u) = \frac{\partial^d}{\partial u_1 \ldots \partial u_d} \mathbf{P}\left(V = v, U \leq u\right), v \in A, u \in \mathbf{R}^d. \tag{16}$$

and the marginal pdf (pdf of U) is $g_U(u) = \sum_{v \in A} g(v, u), u \in \mathbf{R}^d$.

Lemma 14 *In the situation above,*

$$\mathbf{P}\left(V = a|U\right) = \frac{g(a, U)}{\sum_{v \in A} g(v, U)}. \tag{17}$$

(convention $\frac{c}{0} = 0$ applies).

Proof. For any bounded function $h(u), u \in \mathbf{R}^d$,

$$\begin{aligned} \mathbf{E}[1_{V=a} h(U)] &= \int_{\mathbf{R}^d} h(u) g(a,u)\ du \\ &= \int_{\mathbf{R}^d} h(u) \frac{g(a,u)}{g_U(u)}\ g_Y(u) du \\ &= \mathbf{E}[h(U) \frac{g(a,U)}{g_U(U)}\ ]. \end{aligned}$$

■

3.2 Past and future of a Markov sequence

Lemma 15 *Assume Z_n is a sequence of r.v. such that*

$$\mathbf{P}\left(Z_{[n+1,m]} = z_{[n+1,m]} | Z_{[0,n]} = z_{[0,n]}\right) = \mathbf{P}\left(Z_{[n+1,m]} = z_{[n+1,m]} | Z_n = z_n\right).$$

Then

$$\mathbf{P}\left(Z_{[0,n]} = z_{[0,n]} | Z_{[n+1,m]} = z_{[n+1,m]}\right) = \mathbf{P}\left(Z_{[0,n]} = z_{[0,n]} | Z_{n+1} = z_{n+1}\right). \quad (18)$$

Proof. Indeed,

$$\begin{aligned} LHS &= \frac{\mathbf{P}\left(Z_{[0,n]} = z_{[0,n]}, Z_{[n+1,m]} = z_{[n+1,m]}\right)}{\mathbf{P}\left(Z_{[n+1,m]} = z_{[n+1,m]}\right)} \\ &= \frac{\mathbf{P}\left(Z_{[n+2,m]} = z_{[n+2,m]} | Z_{[0,n+1]} = z_{[0,n+1]}\right) \mathbf{P}\left(Z_{[0,n+1]} = z_{[0,n+1]}\right)}{\mathbf{P}\left(Z_{[n+2,m]} = z_{[n+2,m]} | Z_{n+1} = z_{n+1}\right) \mathbf{P}\left(Z_{n+1} = z_{n+1}\right)} \\ &= RHS. \end{aligned}$$

■

Note that the properties above imply that for any bounded function $f(z_{[n+1,T]})$

$$\mathbf{E}[f\left(Z_{[n+1,T]}\right)|Z_{[0,n]}] = \mathbf{E}[f\left(Z_{[n+1,T]}\right)|Z_n],$$

and for any bounded $f(z_{[0,n]})$,

$$\mathbf{E}[f\left(Z_{[0,n]}\right)|Z_{[n+1,m]}] = \mathbf{E}[f\left(Z_{[0,n]}\right)|Z_{n+1}].$$

As following statement shows this does not change if we include some independent r.v.

Lemma 16 *Assume Z_n is a homogeneous Markov chain and U, V are independent r.v. (U, V, Z are independent). Then*

$$\begin{aligned} &\mathbf{E}\left[f\left(Z_{[n+1,T]}, V\right)|Z_{[0.n]}, U)\right] \\ =\ &\mathbf{E}\left[f\left(Z_{[n+1,T]}, V\right)|Z_{[0.n]}\right] \\ =\ &\mathbf{E}\left[f\left(Z_{[n+1,T]}, V\right)|Z_n\right]. \end{aligned}$$

Similarly,

$$\begin{aligned} & \mathbf{E}\left[f\left(Z_{[0,n]}, V\right)|Z_{[n+1,T]}, U\right)\right] \\ = \ & \mathbf{E}\left[f\left(Z_{[0,n]}, V\right)|Z_{[n+1,T]}\right] \\ = \ & \mathbf{E}\left[f\left(Z_{[0,n]}, V\right)|Z_{n+1}\right]. \end{aligned}$$

Proof. By Lemma 15, for any bounded function $g(x_{[0,n]}, u)$,

$$\begin{aligned} \mathbf{E}\left[f\left(Z_{[n+1,T]}, V\right) g(Z_{[0.n]}, U)\right] &= \mathbf{E}\left\{\mathbf{E}f(c', V)|_{c'=Z_{[n+1,T]}} \mathbf{E}g(c, U)|_{c=Z_{[0,n]}}\right\} \\ & \quad \mathbf{E}\left\{\mathbf{E}[\mathbf{E}f(c', V)|_{c'=Z_{[n+1,T]}}|Z_n]\mathbf{E}g(c, U)|_{c=Z_{[0,n]}}\right\} \\ &= \mathbf{E}\left\{\mathbf{E}[f\left(Z_{[n+1,T]}, V\right)|Z_n]\mathbf{E}g(c, U)|_{c=Z_{[0,n]}}\right\} \\ &= \mathbf{E}\left\{\mathbf{E}[f\left(Z_{[n+1,T]}, V\right)|Z_n]g(Z_{[0,n]}, U)\right\}. \end{aligned}$$

Similarly, for any bounded function $g(x_{[n+1,T]}, u)$,

$$\begin{aligned} \mathbf{E}\left[f\left(Z_{[0,n]}, V\right) g(Z_{[n+1,T]}, U)\right] &= \mathbf{E}\left\{\mathbf{E}f(c', V)|_{c'=Z_{[0,n,]}} \mathbf{E}g(c, U)|_{c=Z_{[n+1,T]}}\right\} \\ &= \mathbf{E}\left\{\mathbf{E}[f\left(Z_{[0,n]}, V\right)|Z_{n+1}]\mathbf{E}g(c, U)|_{c=Z_{[n+1,T]}}\right\} \\ &= \mathbf{E}\left\{\mathbf{E}[f\left(Z_{[n+1,T]}, V\right)|Z_{n+1}]g(Z_{[n+1,T]}, U)\right\}. \end{aligned}$$

■