Derivations for Temporal Models

For those who prefer a more formal treatment, below are formal derivations for the recursive formulas given in class for filtering, prediction, smoothing and finding the most likely sequence. R&N also provides such derivations, but the ones given here are meant to go along more closely with the way that I did things in class.

Filtering

We want to compute $P(x_t|\mathbf{e}_{1:t})$. Note that, by definition of conditional probability,

$$P(x_t|\mathbf{e}_{1:t}) = \frac{P(x_t, \mathbf{e}_{1:t})}{P(\mathbf{e}_{1:t})}$$

so $P(x_t|\mathbf{e}_{1:t}) \propto P(x_t, \mathbf{e}_{1:t})$ for any t.

We derive a recursive expression as follows:

$$\begin{aligned}
P(x_{t+1}|\mathbf{e}_{1:t+1}) &\propto P(x_{t+1}, \mathbf{e}_{1:t+1}) \\
&= \sum_{x_t} P(x_t, x_{t+1}, \mathbf{e}_{1:t+1}) && \text{marginalization} \\
&= \sum_{x_t} P(x_t, \mathbf{e}_{1:t}, x_{t+1}, e_{t+1}) && \text{breaking } \mathbf{e}_{1:t+1} \text{ into } \mathbf{e}_{1:t} \text{ and } e_{t+1} \\
&= \sum_{x_t} P(x_t, \mathbf{e}_{1:t}) P(x_{t+1}, e_{t+1}|x_t, \mathbf{e}_{1:t}) && \text{definition of conditional probability} \\
&= \sum_{x_t} P(x_t, \mathbf{e}_{1:t}) P(x_{t+1}|x_t, \mathbf{e}_{1:t}) P(e_{t+1}|x_{t+1}, x_t, \mathbf{e}_{1:t}) && \text{definition of conditional probability} \\
&= \sum_{x_t} P(x_t, \mathbf{e}_{1:t}) P(x_{t+1}|x_t) P(e_{t+1}|x_{t+1}) && \text{by the Markov assumptions (applied twice)} \\
&= P(e_{t+1}|x_{t+1}) \sum_{x_t} P(x_t, \mathbf{e}_{1:t}) P(x_{t+1}|x_t) && \text{factoring out a constant from the sum} \\
&\propto P(e_{t+1}|x_{t+1}) \sum_{x_t} P(x_t|\mathbf{e}_{1:t}) P(x_{t+1}|x_t) && \text{by the comments above.}
\end{aligned}$$

Prediction

We want to compute $P(x_{t+k}|\mathbf{e}_{1:t})$. We again derive a recursive expression:

$$\begin{aligned}
P(x_{t+k+1}|\mathbf{e}_{1:t}) &= \sum_{x_{t+k}} P(x_{t+k}, x_{t+k+1}|\mathbf{e}_{1:t}) && \text{using marginalization} \\
&= \sum_{x_{t+k}} P(x_{t+k}|\mathbf{e}_{1:t}) P(x_{t+k+1}|x_{t+k}, \mathbf{e}_{1:t}) && \text{definition of conditional probability} \\
&= \sum_{x_{t+k}} P(x_{t+k}|\mathbf{e}_{1:t}) P(x_{t+k+1}|x_{t+k}) && \text{by the Markov assumptions.}
\end{aligned}$$

Smoothing

We want to compute $P(x_k|\mathbf{e}_{1:t})$, for $k < t$. We have:

$$\begin{aligned}
P(x_k|\mathbf{e}_{1:t}) &\propto P(x_k, \mathbf{e}_{1:t}) && \text{by the usual argument} \\
&= P(x_k, \mathbf{e}_{1:k}, \mathbf{e}_{k+1:t}) && \text{breaking up } \mathbf{e}_{1:t} \text{ into } \mathbf{e}_{1:k} \text{ and } \mathbf{e}_{k+1:t} \\
&= P(x_k, \mathbf{e}_{1:k})\, P(\mathbf{e}_{k+1:t}|x_k, \mathbf{e}_{1:k}) && \text{definition of conditional probability} \\
&= P(x_k, \mathbf{e}_{1:k})\, P(\mathbf{e}_{k+1:t}|x_k) && \text{by the Markov assumptions} \\
&\propto P(x_k|\mathbf{e}_{1:k})\, P(\mathbf{e}_{k+1:t}|x_k).
\end{aligned}$$

We already saw how to compute $P(x_k|\mathbf{e}_{1:k})$. For the other factor, we can do a recursive computation:

$$\begin{aligned}
P(\mathbf{e}_{k+1:t}|x_k) &= \sum_{x_{k+1}} P(x_{k+1}, \mathbf{e}_{k+1:t}|x_k) && \text{marginalization} \\
&= \sum_{x_{k+1}} P(x_{k+1}|x_k)\, P(\mathbf{e}_{k+1:t}|x_k, x_{k+1}) && \text{definition of conditional probability} \\
&= \sum_{x_{k+1}} P(x_{k+1}|x_k)\, P(\mathbf{e}_{k+1:t}|x_{k+1}) && \text{by the Markov assumptions} \\
&= \sum_{x_{k+1}} P(x_{k+1}|x_k)\, P(e_{k+1}, \mathbf{e}_{k+2:t}|x_{k+1}) && \text{breaking up } \mathbf{e}_{k+1:t} \\
&= \sum_{x_{k+1}} P(x_{k+1}|x_k)\, P(e_{k+1}|x_{k+1})\, P(\mathbf{e}_{k+2:t}|e_{k+1}, x_{k+1}) && \text{definition of conditional probability} \\
&= \sum_{x_{k+1}} P(x_{k+1}|x_k)\, P(e_{k+1}|x_{k+1})\, P(\mathbf{e}_{k+2:t}|x_{k+1}) && \text{by the Markov assumptions.}
\end{aligned}$$

Finding the most likely sequence

We wish to find the state sequence $\mathbf{x}_{0:t}$ that maximizes $P(\mathbf{x}_{0:t}|\mathbf{e}_{1:t})$. Since they only differ by a constant factor, this is the same as maximizing $P(\mathbf{x}_{0:t}, \mathbf{e}_{1:t})$. It is enough, for all x_t, to find the maximum over $\mathbf{x}_{0:t-1}$, since then, as a final step, we can take a final maximum over x_t. In other words, we can use the fact that

$$\max_{\mathbf{x}_{0:t}} P(\mathbf{x}_{0:t}, \mathbf{e}_{1:t}) = \max_{x_t} \left[\max_{\mathbf{x}_{0:t-1}} P(\mathbf{x}_{0:t}, \mathbf{e}_{1:t}) \right].$$

As usual, we will derive a recursive expression:

$$
\begin{aligned}
&\max_{\mathbf{x}_{0:t-1}} P\left(\mathbf{x}_{0:t}, \mathbf{e}_{1:t}\right) \\
&\quad = \max_{\mathbf{x}_{0:t-1}} P\left(\mathbf{x}_{0:t-1}, x_t, \mathbf{e}_{1:t-1}, e_t\right) && \text{breaking up } \mathbf{x}_{0:t} \text{ and } \mathbf{e}_{1:t} \\
&\quad = \max_{\mathbf{x}_{0:t-1}} \left[P\left(\mathbf{x}_{0:t-1}, \mathbf{e}_{1:t-1}\right) P\left(x_t | \mathbf{x}_{0:t-1}, \mathbf{e}_{1:t-1}\right) P\left(e_t | x_t, \mathbf{x}_{0:t-1}, \mathbf{e}_{1:t-1}\right)\right] && \text{definition of conditional probability (applied repeatedly)} \\
&\quad = \max_{\mathbf{x}_{0:t-1}} \left[P\left(\mathbf{x}_{0:t-1}, \mathbf{e}_{1:t-1}\right) P\left(x_t | x_{t-1}\right) P\left(e_t | x_t\right)\right] && \text{by the Markov assumptions (applied twice)} \\
&\quad = \max_{x_{t-1}} \max_{\mathbf{x}_{0:t-2}} \left[P\left(\mathbf{x}_{0:t-1}, \mathbf{e}_{1:t-1}\right) P\left(x_t | x_{t-1}\right) P\left(e_t | x_t\right)\right] && \text{breaking up the maximum} \\
&\quad = \max_{x_{t-1}} \left[P\left(x_t | x_{t-1}\right) P\left(e_t | x_t\right) \max_{\mathbf{x}_{0:t-2}} P\left(\mathbf{x}_{0:t-1}, \mathbf{e}_{1:t-1}\right)\right] && \text{factoring out constant terms from the inner maximum.}
\end{aligned}
$$

Note that in the base case, $t = 0$, we have

$$\max_{\mathbf{x}_{0:t-1}} P\left(\mathbf{x}_{0:t}, \mathbf{e}_{1:t}\right) = P\left(x_0\right).$$