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 $P(x_t|e_{1:t})$. Note that, by definition of conditional probability,

$$P(x_t|e_{1:t}) = \frac{P(x_t, e_{1:t})}{P(e_{1:t})}$$

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

We derive a recursive expression as follows:

$$P(x_{t+1}|e_{1:t+1}) \propto P(x_{t+1}, e_{1:t+1})$$

$$= \sum_{x_t} P(x_t, x_{t+1}, e_{1:t+1})$$

$$= \sum_{x_t} P(x_t, e_{1:t}, x_{t+1}, e_{t+1})$$

$$= \sum_{x_t} P(x_t, e_{1:t}) P(x_{t+1}, e_{t+1}|x_t, e_{1:t})$$

$$= \sum_{x_t} P(x_t, e_{1:t}) P(x_{t+1}|x_t, e_{1:t}) P(e_{t+1}|x_{t+1}, x_t, e_{1:t})$$

$$= \sum_{x_t} P(x_t, e_{1:t}) P(x_{t+1}|x_t) P(e_{t+1}|x_{t+1})$$

$$= P(e_{t+1}|x_{t+1}) \sum_{x_t} P(x_t, e_{1:t}) P(x_{t+1}|x_t)$$

$$\propto P(e_{t+1}|x_{t+1}) \sum_{x_t} P(x_t|e_{1:t}) P(x_{t+1}|x_t)$$

Thus, $P(x_{t+1}|e_{1:t+1})$ can be computed recursively from $P(x_t|e_{1:t})$. In the base case that $t = 0$, we use $P(x_0|e_{1:0}) = P(x_0)$.

**Prediction**

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

$$P(x_{t+k+1}|e_{1:t}) = \sum_{x_{t+k}} P(x_{t+k}, x_{t+k+1}|e_{1:t})$$

using marginalization

$$= \sum_{x_{t+k}} P(x_{t+k}|e_{1:t}) P(x_{t+k+1}|x_{t+k}, e_{1:t})$$

definition of conditional probability

$$= \sum_{x_{t+k}} P(x_{t+k}|e_{1:t}) P(x_{t+k+1}|x_{t+k})$$

by the Markov assumptions.

In the base case that $k = 0$, we compute $P(x_t|e_{1:t})$ using the filtering algorithm above.
Smoothing

We want to compute $P(x_k|e_1:t)$, for $k < t$. We have:

$$P(x_k|e_1:t) \propto P(x_k, e_1:t)$$

by the usual argument

$$= P(x_k, e_{1:k}, e_{k+1:t})$$

breaking up $e_{1:t}$ into $e_{1:k}$ and $e_{k+1:t}$

$$= P(x_k, e_{1:k}) P(e_{k+1:t}|x_k, e_{1:k})$$

definition of conditional probability

$$= P(x_k, e_{1:k}) P(e_{k+1:t}|x_k)$$

by the Markov assumptions

$$\propto P(x_k|e_{1:k}) P(e_{k+1:t}|x_k).$$

We already saw how to compute $P(x_k|e_{1:k})$ using the filtering algorithm above. For the other factor $P(e_{k+1:t}|x_k)$, we can

do a (backwards) recursive computation:

$$P(e_{k+1:t}|x_k) = \sum_{x_{k+1}} P(x_{k+1}, e_{k+1:t}|x_k)$$

marginalization

$$= \sum_{x_{k+1}} P(x_{k+1}|x_k) P(e_{k+1:t}|x_k, x_{k+1})$$

definition of conditional probability

$$= \sum_{x_{k+1}} P(x_{k+1}|x_k) P(e_{k+1:t}|x_{k+1})$$

by the Markov assumptions

$$= \sum_{x_{k+1}} P(x_{k+1}|x_k) P(e_{k+1}, e_{k+2:t}|x_{k+1})$$

breaking up $e_{k+1:t}$

$$= \sum_{x_{k+1}} P(x_{k+1}|x_k) P(e_{k+1}|x_{k+1}) P(e_{k+2:t}|e_{k+1}, x_{k+1})$$

definition of conditional probability

$$= \sum_{x_{k+1}} P(x_{k+1}|x_k) P(e_{k+1}|x_{k+1}) P(e_{k+2:t}|x_{k+1})$$

by the Markov assumptions.

In the base case that $k = t$, we use $P(e_{t+1:t}|x_t) = 1$. 
Finding the most likely sequence

(Note that the derivation below corrects the treatment in R&N which erroneously ignores $x_0$.)

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

$$\max_{x_{0:t}} P(x_{0:t}, e_{1:t}) = \max_{x_{0:t-1}} \left[ \max_{x_t} P(x_{0:t}, e_{1:t}) \right].$$

As usual, we will derive a recursive expression:

$$\max_{x_{0:t}} P(x_{0:t}, e_{1:t})$$

$$= \max_{x_{0:t-1}} P(x_{0:t-1}, x_t, e_{1:t-1}, e_t)$$

$$= \max_{x_{0:t-1}} \left[ P(x_{0:t-1}, e_{1:t-1}) P(x_t|x_{0:t-1}, e_{1:t-1}) P(e_t|x_t, x_{0:t-1}, e_{1:t-1}) \right]$$

$$= \max_{x_{0:t-1}} \left[ P(x_{0:t-1}, e_{1:t-1}) P(x_t|x_{t-1}) P(e_t|x_t) \right]$$

$$= \max_{x_{t-1}} \left[ P(x_t|x_{t-1}) P(e_t|x_t) \max_{x_{0:t-2}} P(x_{0:t-1}, e_{1:t-1}) \right]$$

by the Markov assumptions (applied twice)

breaking up $x_{0:t}$ and $e_{1:t}$

factoring out constant terms from the inner maximum.

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

$$\max_{x_{0:t}} P(x_{0:t}, e_{1:t}) = P(x_0).$$