Chapter 13

Derin's Algorithm

13.1 Introduction

Any mathematically precise treatment of the scoring, alignment, and training problems of HMM as stated in the preceding requires a few auxiliary factorizations and conditional independence statements inherent in a standard HMM to be made explicit. These results are consequences of the Markov property and the postulated conditional independence of emission variables given a sequence of states of the hidden Markov chain.

We exploit these properties to prove *Derin's backwards recursive formula for smoothing posterior probabilities* (Askar and Derin 1981), which is a first instance of an algorithmic probability connected with HMM to be presented in this text. Additional such schemes will be based on the forward and backward variables established in this chapter. These assertions about conditional distributions yield, also, sufficient conditions for a function of a Markov chain to be a Markov chain. The results will also be needed in the analysis of consistency of the maximum likelihood estimate for HMM in Chapter 17 below. The present chapter has a technical nature. The statements are in most cases intuitively obvious, and can be visualized by suitably separating the graph in Figure 13.1 but some of the proofs, if carried out in detail, are laborious and can be omitted at a first reading.

The main trick in the computations to follow is to calculate a probability by including suitable extra random variables in the argument of a probability so as to obtain a desired probability by marginalization. In this chapter we shall follow (Lucke 1996) and use the notation $\bigotimes$ in several of the marginal-

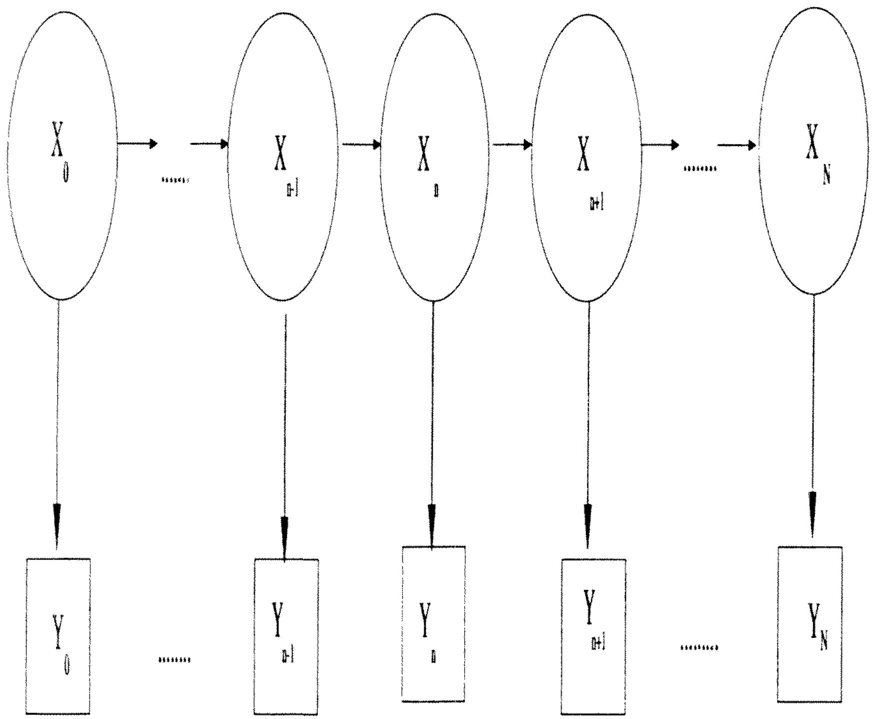

Figure 13.1: HMM: Influence diagram

izations when these require summations over a set of variables. In addition, the chain rule (2.16) in Chapter 1 is used in various extended forms, sometimes without explicit reference.

13.2 Derin's Formula

The goal is to find a backwards recursion with $n \leq m$ for the *smoothing posterior probability* defined as

$$\widehat{\pi}_j (n|m) \doteq P (X_n = j|Y_0 = o_0, \ldots, Y_m = o_m) \tag{2.1}$$

for a standard HMM. The result below is found in (Askar and Derin 1981) and seems to be known as Derin's formula (see, e.g., Devijver 1985).

Proposition 13.2.1 *For $n = 0, \ldots, N-1$*

$$\widehat{\pi}_j (n|N) = \widehat{\pi}_j (n|n) \cdot \sum_{k=1}^{J} \frac{a_{j|k}}{\widehat{\pi}_k (n+1|n)} \widehat{\pi}_k (n+1|N) . \tag{2.2}$$

■

This proposition will be proved in Section 13.7 using the results on conditional distributions established in the sections preceding Section 13.7. A forward recursion for the *filtering posterior probability* $\widehat{\pi}_j (n|n)$ is given in Exercise 3 in equation (8.1) and a forward recursion for the *prediction posterior probability* $\widehat{\pi}_k (n+1|n)$ in Exercise 3 in (8.2).

Thus there is a computational scheme in (2.2): $\widehat{\pi}_j (n|n)$ and $\widehat{\pi}_k (n+1|n)$ are computed recursively forwards and then stored to compute $\widehat{\pi}_j (n|N)$ recursively backwards. This has also bearing on a numerical implementation issue to be made explicit later. Derin's algorithm works with quantities normalized at each step that eliminate the underflow problem inherent in Baum's forward–backward procedure for computation of likelihoods.

In (Guédon and Cocozza-Thivent 1990) it is shown that Derin's formula can be extended to derive a learning algorithm for HMM with duration (hidden semi-Markov Model).

13.3 Various Sufficiency Properties

Here the typesetting is simplified, e.g., by writing a conditional probability

$$P (Y_m = o_m, \ldots, Y_N = o_N \mid X_n = j_n \ldots, X_N = j_N)$$

simply as

$$P (Y_m, \ldots, Y_N \mid X_n, \ldots, X_N)$$

without a large risk of confusion in the proofs to follow. Thus conditional independence becomes

$$P (Y_0, \ldots, Y_n \mid X_0, \ldots, X_n) = \prod_{i=0}^{n} P (Y_i \mid X_i) . \tag{3.1}$$

The next two propositions in this section have the common feature of showing that the probability of a finite length emitted sequence conditioned on a state sequence of the hidden Markov chain depends only on a subsequence of the state sequence. The relevant subsequence can be seen as a sufficient reduction of the hidden Markov chain. The general idea of sufficiency is defined and described in exercises to Chapter 4 above. The properties are implicit in a statement of some five lines' length in the basic work (Baum et al. 1970) and are made explicitly available in (MacDonald and Zucchini 1997). Precautions with regard to some conditioning event having zero probability are not explicitly taken. In those cases both sides of the expressions will be taken as zero.

Proposition 13.3.1 *For all integers n and m such that $0 \leq n \leq m \leq N$*

$$P(Y_m, \ldots, Y_N \mid X_n \ldots, X_N) = P(Y_m, \ldots, Y_N \mid X_m, \ldots, X_N). \tag{3.2}$$

Proof: The left hand side of (3.2) can be expressed as

$$\frac{1}{P(X_n, \ldots, X_N)} \bigotimes P(Y_m, \ldots, Y_N \mid X_0 \ldots, X_N) \cdot P(X_0 \ldots, X_N),$$

where $\bigotimes$ designates the summation over $j_0, \ldots, j_{n-1}$ (which are the values of $X_0, \ldots, X_{n-1}$). If $n = 0$ there is no summation. By assumption (3.1) (and a marginalization argument)

$$P(Y_m, \ldots, Y_N \mid X_0 \ldots, X_N) = P(Y_m|X_m) \cdot \ldots \cdot P(Y_N|X_N).$$

This factor can be taken outside the summation sign $\bigotimes$, since $m \geq n$. This leaves us with

$$\prod_{l=m}^{N} P(Y_l|X_l) \frac{1}{P(X_n, \ldots, X_N)} \bigotimes P(X_0 \ldots, X_N),$$

where the sum is equal to $P(X_n, \ldots, X_N)$, since we are summing over $j_0, \ldots, j_{n-1}$. Thus the whole last expression is equal to

$$= \prod_{l=m}^{N} P(Y_l|X_l),$$

which is independent of n. Since the right hand side of (3.2) is a special case of the left hand side for $n = m$, this proves the assertion as claimed. ∎

We shall also in the rest of the proofs write for convenience of expression

$$X^{(t)} = (X_0, \ldots, X_t), Y^{(t)} = (Y_0, \ldots, Y_t),$$

where we use again the convention of dropping the $X_n = j_n$'s and $Y_n = o_n$'s.

Proposition 13.3.2 *For all integers $n = 0, \ldots, N-1$*

$$P(Y_{n+1}, \ldots, Y_N \mid X_0, \ldots, X_n) = P(Y_{n+1}, \ldots, Y_N \mid X_n). \tag{3.3}$$

Proof: The left hand side of (3.3) is

$$\frac{1}{P(X^{(n)})} \bigotimes P(X^{(N)}) \cdot P(Y_{n+1}, \ldots, Y_N \mid X^{(N)}), \tag{3.4}$$

where the summation $\bigotimes$ is over $j_{n+1}, \ldots, j_N$. By Proposition 13.3.1, equation (3.2) (with $m = n+1$, $n = 0$), we have

$$P(Y_{n+1}, \ldots, Y_N \mid X^{(N)}) = P(Y_{n+1}, \ldots, Y_N \mid X_{n+1}, \ldots, X_N),$$

and using the same proposition and equation once more (with $m = n+1$) we have

$$P(Y_{n+1}, \ldots, Y_N \mid X_{n+1}, \ldots, X_N) = P(Y_{n+1}, \ldots, Y_N \mid X_n, \ldots, X_N).$$

Thus we are dealing with

$$\bigotimes P(X^{(N)}) \cdot P(Y_{n+1}, \ldots, Y_N \mid X^{(N)})$$
$$= \bigotimes P(X^{(N)}) P(Y_{n+1}, \ldots, Y_N \mid X_n, \ldots, X_N).$$

By conditional probability $P(X^{(N)}) = P(X_{n+1}, \ldots, X_N|X^{(n)}) \cdot P(X^{(n)})$ and by a consequence of *Markov property* (see Exercise 5.5.1 (c) in Chapter 7) we have

$$P(X_{n+1}, \ldots, X_N|X^{(n)}) = P(X_{n+1}, \ldots, X_N|X_n).$$

Thus the sum is, by an extension of the chain rule (2.16) in Chapter 1, equal to

$$\bigotimes P(X^{(N)}) \cdot P(Y_{n+1}, \ldots, Y_N \mid X_n, \ldots, X_N)$$
$$= \bigotimes \frac{P(Y_{n+1}, \ldots, Y_N, X_n, \ldots, X_N) \cdot P(X^{(n)})}{P(X_n)}, \tag{3.5}$$

and since we are in $\bigotimes$ summing over $j_{n+1}, \ldots, j_N$ we obtain this as

$$P(X^{(n)}) \bigotimes \frac{P(Y_{n+1}, \ldots, Y_N, X_n, \ldots, X_N)}{P(X_n)}$$

and

$$\bigotimes \frac{P(Y_{n+1},\ldots,Y_N,X_n,\ldots,X_N)}{P(X_n)} = \frac{P(Y_{n+1},\ldots,Y_N,X_n)}{P(X_n)} = P(Y_{n+1},\ldots,Y_N|X_n).$$

Thus in (3.4) we have found that

$$\frac{1}{P(X^{(n)})}\bigotimes P\left(X^{(N)}\right)\cdot P\left(Y_{n+1},\ldots,Y_N \mid X^{(N)}\right) = \frac{1}{P(X^{(n)})}P(Y_{n+1},\ldots,Y_N|X_n)\cdot P\left(X^{(n)}\right),$$

which proves the assertion. ∎

Proposition 13.3.3 *For all integers $n = 0,\ldots,N$*

$$P(Y_0,\ldots,Y_n \mid X_0,\ldots,X_N) = P(Y_0,\ldots,Y_n \mid X_0,\ldots,X_n). \tag{3.6}$$

Proof: The left hand side of (3.6) is, with our notational conventions, written as

$$P\left(Y_0,\ldots,Y_n \mid X^{(N)}\right) = \bigotimes P\left(Y_0,\ldots,Y_N \mid X^{(N)}\right),$$

where the sum $\bigotimes$ is over $o_{n+1}, o_{n+2},\ldots,o_N$. But then the conditional independence assumption (3.1) gives

$$\bigotimes P\left(Y_0,\ldots,Y_N \mid X^{(N)}\right) = \prod_{l=0}^{n} P(Y_l|X_l)\bigotimes \prod_{l=n+1}^{N} P(Y_l|X_l)$$

where the summation gives $\bigotimes\prod_{l=n+1}^{N} P(Y_l|X_l) = 1$ or

$$P\left(Y_0,\ldots,Y_n \mid X^{(N)}\right) = \prod_{l=0}^{n} P(Y_l|X_l),$$

as was to be proved. ∎

13.4 Lumping of Markov Chains

It was observed in an exercise of Chapter 10 that the emitted sequence of an HMM can be seen as a function of a Markov chain. Next, using a result obtained in the preceding section, we study the question of whether a function of a Markov chain is another Markov chain.

Now let g be a map from S onto the alphabet $\mathcal{O} = \{o_1,\ldots,o_K\}$. The process $\{Y_n\}_{n=0}^{\infty}$ is defined by

$$Y_n = g(X_n). \tag{4.7}$$

If $g(\cdot)$ is not one-to-one, i.e.,$K < J$, we are obviously lumping (or aggregating) the state space of the Markov chain. Lumping of the state space is often of interest in probabilistic modelling of DNA sequences, for example, $\{A,T,C,G\}$ can be lumped into the purine-pyriminide groups $\{A/G, C/T\}$ (Raftery and Tavaré 1994).

Simple examples show that $\{Y_n\}_{n=0}^{\infty}$ is not, in general, a Markov chain. There has been a considerable research in the characterization of those functions of Markov chains that are themselves Markov chains. The following proposition (Kelly 1982) gives a straightforward and natural set of sufficient conditions for functions of Markov chains to be Markovian functions. An extra process

$$Z_n = \phi(X_n) \tag{4.8}$$

will play the role of a kind of *sufficient statistic* in the sense that it is assumed to contain all the information in X_n relevant to the next state of the process Y_n.

Proposition 13.4.1 *Let $\{X_n\}_{n=0}^{\infty} \in$ Markov (P, p_{X_0}). If*

$$P(Y_{n+1} = o_{n+1} \mid X_n = j) = P(Y_{n+1} = o_{n+1} \mid Z_n = \phi(j)) \tag{4.9}$$

and

$$P(Z_n = z_n \mid Y_n,\ldots,Y_0) = P(Z_n = z_n \mid Y_n), \tag{4.10}$$

then $\{Y_n = g(X_n)\}_{n=0}^{\infty}$ is a Markov chain.

Proof: The Markov property will be verified. By marginalization and the chain rule of conditional probability we get

$$P(Y_{n+1} \mid Y_n,\ldots,Y_0) = \sum P(Y_{n+1}, Z_n = z_n \mid Y_n,\ldots,Y_0)$$

$$= \sum_{z_n} P(Y_{n+1} \mid Z_n = z_n, Y_n, \ldots, Y_0) \cdot P(Z_n = z_n \mid Y_n, \ldots, Y_0).$$

Let us again use marginalization, to the effect that

$$P(Y_{n+1} \mid Z_n = z_n, Y_n, \ldots, Y_0) = \bigotimes P(Y_{n+1}, X_n, \ldots, X_0 \mid Z_n = z_n, Y_n, \ldots, Y_0) \quad (4.11)$$

where the summation $\bigotimes$ is over $j_0, \ldots, j_n$ (the values of $X_0, \ldots, X_n$). Then the right hand side is equal to

$$\bigotimes P(Y_{n+1}|X_n, \ldots, X_0, Z_n = z_n, Y_n, \ldots, Y_0)\, P(X_n, \ldots, X_0 \mid Z_n = z_n, Y_n, \ldots, Y_0).$$

But since $Z_n = z_n, Y_n, \ldots, Y_0$ are, in view of (4.7) and (4.8), no longer random but exactly computable given $X_n, \ldots, X_0$, we have

$$P(Y_{n+1}|X_n, \ldots, X_0, Z_n = z_n, Y_n, \ldots, Y_0) = P(Y_{n+1}|X_n, \ldots, X_0).$$

Next we note that a function of a Markov chain satisfies the conditional independence property (3.1) in a trivial manner, as both sides of the equality are simultaneously either one or zero. Thus we are permitted to use (3.3) to obtain

$$P(Y_{n+1}|X_n, \ldots, X_0) = P(Y_{n+1}|X_n)$$

Then we use (4.9), so that

$$P(Y_{n+1}|X_n) = P(Y_{n+1} \mid Z_n = z_n).$$

But then

$$\bigotimes P(Y_{n+1}|X_n, \ldots, X_0, Z_n = z_n, Y_n, \ldots, Y_0)\, P(X_n, \ldots, X_0 \mid Z_n = z_n, Y_n, \ldots, Y_0)$$

$$= P(Y_{n+1}|Z_n = z_n,) \cdot \bigotimes P(X_n, \ldots, X_0 \mid Z_n = z_n, Y_n, \ldots, Y_0)$$
$$= P(Y_{n+1}|Z_n = z_n,).$$

Thus we have obtained that

$$P(Y_{n+1} \mid Z_n = z_n, Y_n \ldots, Y_0) = P(Y_{n+1} \mid Z_n = z_n) = P(Y_{n+1} \mid Z_n = z_n, Y_n).$$

Using this and the condition (4.10) it follows that

$$P(Y_{n+1} \mid Y_n, \ldots, Y_0) = \sum_{z_n} P(Y_{n+1} \mid Z_n = z_n, Y_n) \cdot P(Z_n = z_n \mid Y_n) = P(Y_{n+1} \mid Y_n),$$

which was to be proved. ∎

If Y_n and Z_n are functions of each other, $Y_n \equiv Z_n$, then (4.9) and (4.10) are reduced to one simple sufficient condition

$$P(Y_{n+1} = o_{n+1} \mid X_n = j) = P(Y_{n+1} = o_{n+1} \mid Y_n = g(j)). \quad (4.12)$$

13.5 Factorizations

The statements in this section are factorizations which demonstrate a kind of renewal structure for the HMM. For example, Proposition 13.5.1 below shows that if at some instants we actually were able to see the state of the hidden chain, then the past and future of the emitted process are independent, conditioned on this one value of the Markov chain. Most of the proofs are based on (MacDonald and Zucchini 1997).

Proposition 13.5.1 *For all integers* $n = 0, \ldots, N$

$$P(Y_0, Y_1, \ldots, Y_N \mid X_n) = P(Y_0, Y_1, \ldots, Y_n \mid X_n) \cdot P(Y_{n+1}, \ldots, Y_N \mid X_n). \quad (5.1)$$

Proof: The left hand side of (5.1) is, by definition of conditional expectation, equal to

$$P\left(Y^{(N)} \mid X_n\right) = \frac{P\left(Y^{(N)}, X_n\right)}{P(X_n)}.$$

Here we again obtain by marginalization that

$$P\left(Y^{(N)}, X_n\right) = \bigotimes_1 \bigotimes_2 P(Y_0, \ldots, Y_N, X_0, \ldots, X_{n-1}, X_n, X_{n+1}, \ldots, X_N), \quad (5.2)$$

where $\bigotimes_1$ is a summation over $j_0, \ldots, j_{n-1}$ and $\bigotimes_2$ is in its turn a summation over $j_{n+1}, \ldots, j_N$. The definition of conditional probability tells that

$$P(Y_0, \ldots, Y_N, X_0, \ldots, X_{n-1}, X_n, X_{n+1}, \ldots, X_N)$$

$$= P\left(Y_0,\ldots,Y_N|X^{(N)}\right)\cdot P\left(X^{(N)}\right)$$

and by the basic assumption (3.1) we have

$$P\left(Y_0,\ldots,Y_N|X^{(N)}\right) = P\left(Y_0,\ldots,Y_n|X^{(n)}\right)\cdot P\left(Y_{n+1},\ldots,Y_N|X_{n+1},\ldots,X_N\right).$$

Next we obtain from (3.2) in Proposition 13.3.1 (set $m = n+1$) that

$$P\left(Y_{n+1},\ldots,Y_N|X_{n+1},\ldots,X_N\right) = P\left(Y_{n+1},\ldots,Y_N|X_n,X_{n+1},\ldots,X_N\right).$$

By conditional probability and the Markov property, see Exercise 5.5.1 (c) in Chapter 7, it follows that

$$\begin{aligned} P\left(X^{(N)}\right) &= P\left(X^{(n)}\right)P\left(X_{n+1},\ldots,X_N|X^{(n)}\right) \\ &= P\left(X^{(n)}\right)P\left(X_{n+1},\ldots,X_N|X_n\right). \end{aligned}$$

Inserting from the preceding we obtain

$$P\left(Y_0,Y_1,\ldots,Y_N|X^{(N)}\right)\cdot P\left(X^{(N)}\right) = T_1\cdot T_2,$$

where

$$\begin{aligned} T_1 &= P\left(Y_0,\ldots,Y_n|X^{(n)}\right)P\left(X^{(n)}\right) \\ T_2 &= P\left(Y_{n+1},\ldots,Y_N|X_n,X_{n+1},\ldots,X_N\right)P\left(X_{n+1},\ldots,X_N|X_n\right). \end{aligned}$$

Using an extension of the chain rule (2.16) in Chapter 1 we then rewrite T_2 to obtain

$$T_1\cdot T_2 = P\left(Y_0,\ldots,Y_n|X^{(n)}\right)P\left(X^{(n)}\right)\cdot P\left(Y_{n+1},\ldots,Y_N,X_{n+1},\ldots,X_N|X_n\right)$$

Therefore in (5.2) we have, since $\bigotimes_2$ is a summation over $j_{n+1},\ldots,j_N$,

$$\begin{aligned} &\bigotimes_1\bigotimes_2 P\left(Y_0,\ldots,Y_N,X_0,\ldots,X_{n-1},X_n,X_{n+1},\ldots,X_N\right) \\ &= \bigotimes_1 P\left(Y_0,\ldots,Y_n|X^{(n)}\right)P\left(X^{(n)}\right) \\ &\quad \bigotimes_2 P\left(Y_{n+1},\ldots,Y_N,X_{n+1},\ldots,X_N|X_n\right) \\ &= \bigotimes_1 P\left(Y_0,\ldots,Y_n|X^{(n)}\right)\cdot P\left(X^{(n)}\right)\cdot P\left(Y_{n+1},\ldots,Y_N|X_n\right) \end{aligned}$$

and therefore

$$P\left(Y^{(N)}\mid X_n\right) = \frac{\bigotimes_1 P\left(Y_0,\ldots,Y_n,X_0,\ldots,X_n\right)\cdot P\left(Y_{n+1},\ldots,Y_N|X_n\right)}{P\left(X_n\right)}. \quad (5.3)$$

By summing over $j_0,\ldots,j_{n-1}$, in (5.3) we obtain

$$\begin{aligned} &\bigotimes_1 P\left(Y_0,\ldots,Y_n,X_0,\ldots,X_n\right)\cdot P\left(Y_{n+1},\ldots,Y_N|X_n\right) \\ &= P\left(Y_0,\ldots,Y_n,X_n\right)\cdot P\left(Y_{n+1},\ldots,Y_N|X_n\right) \\ &= P\left(Y_0,\ldots,Y_n|X_n\right)\cdot P\left(X_n\right)\cdot P\left(Y_{n+1},\ldots,Y_N|X_n\right), \end{aligned} \quad (5.4)$$

and after division by $P(X_n)$ this gives $P\left(Y^{(N)},X_n\right)/P\left(X_n\right)$ as

$$P\left(Y_0,\ldots,Y_n|X_n\right)\cdot P\left(Y_{n+1},\ldots,Y_N|X_n\right),$$

which is equal to the right hand side of (5.1), as was claimed. ■

The conditional probability $P\left(Y_{n+1},\ldots,Y_N\mid X_n\right)$ will be later called the **backward variable**. The next proposition is used to find a recursion for this backward variable.

Proposition 13.5.2 *For all integers* $n = 0,\ldots,N$

$$P\left(Y_n,Y_{n+1},\ldots,Y_N\mid X_n\right) = P\left(Y_n\mid X_n\right)\cdot P\left(Y_{n+1},\ldots,Y_N\mid X_n\right). \quad (5.5)$$

Proof: This is obtained from (5.1) by summing over $o_0,o_1,\ldots,o_{n-1}$. ■

The following proposition is a variant of Proposition 13.5.1 and is used in the proof of Derin's backwards recursion below.

Proposition 13.5.3 *For all integers* $n = 0,\ldots,N-1$

$$P\left(Y_0,\ldots,Y_N\mid X_{n+1}\right) = P\left(Y_0,\ldots,Y_n\mid X_{n+1}\right)P\left(Y_{n+1},\ldots,Y_N\mid X_{n+1}\right). \quad (5.6)$$

Proof: The proof is a simple modification of the proof of Proposition 13.5.1 above. We start with

$$P\left(Y^{(N)}\mid X_{n+1}\right) = \frac{P\left(Y^{(N)},X_{n+1}\right)}{P\left(X_{n+1}\right)}$$

to obtain by marginalization

$$P\left(Y^{(N)},X_n\right) = \bigotimes_1\bigotimes_2 P\left(Y_0,\ldots,Y_N,X_0,\ldots,X_n,X_{n+1},X_{n+2},\ldots,X_N\right),$$

where $\bigotimes_1$ is now a summation over $j_0, \ldots, j_n$ and $\bigotimes_2$ is a summation over $j_{n+2}, \ldots, j_N$. As in the proof above we obtain

$$P\left(Y_0, Y_1, \ldots, Y_N | X^{(N)}\right) \cdot P\left(X^{(N)}\right) = T_1 \cdot T_2$$

with

$$\begin{aligned} T_1 &= P\left(Y_0, \ldots, Y_n | X^{(n)}\right) P\left(X^{(n+1)}\right) \\ T_2 &= P\left(Y_{n+1}, \ldots, Y_N | X_{n+1}, \ldots, X_N\right) \cdot P\left(X_{n+2}, \ldots, X_N | X_{n+1}\right). \end{aligned}$$

Then

$$\begin{aligned} T_1 \cdot T_2 &= P\left(Y_0, \ldots, Y_n | X^{(n)}\right) P\left(X^{(n+1)}\right) \\ &\quad \cdot P\left(Y_{n+1}, \ldots, Y_N, X_{n+2}, \ldots, X_N | X_{n+1}\right). \end{aligned}$$

Therefore, since $\bigotimes_2$ is a summation over $j_{n+2}, \ldots, j_N$, we have

$$\begin{aligned} &\bigotimes_1 \bigotimes_2 P\left(Y_0, \ldots, Y_N, X_0, \ldots, X_{n-1}, X_n, X_{n+1}, \ldots, X_N\right) \\ &= \bigotimes_1 P\left(Y_0, \ldots, Y_n | X^{(n)}\right) P\left(X^{(n+1)}\right) \cdot \\ &\quad \bigotimes_2 P\left(Y_{n+1}, \ldots, Y_N, X_{n+2}, \ldots, X_N | X_{n+1}\right) \\ &= \bigotimes_1 P\left(Y_0, \ldots, Y_n | X^{(n)}\right) \cdot P\left(X^{(n+1)}\right) \cdot P\left(Y_{n+1}, \ldots, Y_N | X_{n+1}\right). \end{aligned}$$

Owing to (3.6) Proposition 13.3.3, we obtain

$$P\left(Y_0, \ldots, Y_n | X^{(n)}\right) = P\left(Y_0, \ldots, Y_n | X^{(n+1)}\right).$$

Therefore

$$P\left(Y^{(N)} \mid X_{n+1}\right) = \frac{\bigotimes_1 P\left(Y_0, .., Y_n, X_0, .., X_{n+1}\right) P\left(Y_{n+1}, .., Y_N | X_{n+1}\right)}{P\left(X_{n+1}\right)}.$$

The desired result follows by summing over $j_0, \ldots, j_n$ in the expression above as in the proof of Proposition 13.5.1. ∎

The next proposition turns out to be directly instrumental in finding the properties of what will be called **forward variables**.

Proposition 13.5.4 *For all integers $n = 0, \ldots, N-1$*

$$P\left(Y_0, \ldots, Y_N \mid X_n, X_{n+1}\right) = P\left(Y_0, \ldots Y_n \mid X_n\right) \cdot P\left(Y_{n+1}, \ldots, Y_N \mid X_{n+1}\right). \quad (5.7)$$

Proof: We start as before by writing the left hand side of (5.7) as

$$\frac{1}{P\left(X_n, X_{n+1}\right)} \bigotimes_1 \bigotimes_2 P\left(Y_0, Y_1, \ldots, Y_N, X_0, X_1, \ldots, X_n, X_{n+1}, X_{n+2} \ldots, X_N\right),$$

where $\bigotimes_1$ is a summation over $j_0, \ldots, j_n$ and $\bigotimes_2$ is a summation over $j_{n+2}, \ldots, j_N$. Then, as in the proof of Proposition 13.5.1, we have

$$\begin{aligned} &P\left(Y_0, Y_1, \ldots, Y_N, X_0, X_1, \ldots, X_{n-1}, X_n, X_{n+1}, \ldots, X_N\right) \\ &= P\left(X^{(N)}\right) \cdot P\left(Y_0, \ldots, Y_n | X^{(n)}\right) \cdot P\left(Y_{n+1}, \ldots, Y_N | X_{n+1}, \ldots, X_N\right), \end{aligned}$$

where we have used assumption (3.1). By definition of conditional probability we get

$$\begin{aligned} &P\left(X^{(N)}\right) \cdot P\left(Y_0, \ldots, Y_n | X^{(n)}\right) \\ &= P\left(X^{(n)}\right) \cdot P\left(Y_0, \ldots, Y_n | X^{(n)}\right) \cdot P\left(X_{n+1}, \ldots, X_N | X^{(n)}\right). \end{aligned}$$

Inserting these in the double sum we have

$$\bigotimes_1 \bigotimes_2 P\left(Y_0, \ldots, Y_N, X_0, \ldots, X_{n-1}, X_n, X_{n+1}, \ldots, X_N\right) = S_1 \cdot S_2,$$

where

$$\begin{aligned} S_1 &= \bigotimes_1 P\left(X^{(n)}\right) \cdot P\left(Y_0, .., Y_n | X^{(n)}\right) \\ S_2 &= \bigotimes_2 P\left(Y_{n+1}, .., Y_N | X_{n+1}, .., X_N\right) P\left(X_{n+1}, .., X_N | X^{(n)}\right) \end{aligned} \quad (5.8)$$

recalling the nature of the summation $\bigotimes_2$. We evaluate next the sum S_2. We have for the generic term in $\bigotimes_2$

$$\begin{aligned} &P\left(Y_{n+1}, \ldots, Y_N | X_{n+1}, \ldots, X_N\right) \cdot P\left(X_{n+1}, \ldots, X_N | X^{(n)}\right) \\ &= \frac{P\left(Y_{n+1}, \ldots, Y_N, X_{n+1}, \ldots, X_N\right) \cdot P\left(X_{n+1}, \ldots, X_N | X_n\right)}{P\left(X_{n+1}, \ldots, X_N\right)}, \end{aligned}$$

where we have also used the Markov property. Now

$$\frac{P(X_{n+1},\ldots,X_N|X_n)}{P(X_{n+1},\ldots,X_N)} = \frac{P(X_n,X_{n+1},\ldots,X_N)}{P(X_n)\cdot P(X_{n+1},\ldots,X_N)}$$

and the Markov property gives this as

$$\frac{P(X_n)\prod_{l=n}^{N-1}P(X_{l+1}|X_l)}{P(X_n)\,P(X_{n+1})\prod_{l=n+1}^{N-1}P(X_{l+1}|(X_l)}$$

which is clearly equal to

$$\frac{P(X_{n+1}|X_n)}{P(X_{n+1})}.$$

This last ratio is independent of the summation variables in $\bigotimes_2$. Hence

$$\begin{aligned}&\bigotimes_2 P(Y_{n+1},\ldots,Y_N|X_{n+1},\ldots,X_N)\cdot P\left(X_{n+1},\ldots,X_N|X^{(n)}\right)\\ &=\frac{P(X_{n+1}|X_n)}{P(X_{n+1})}\bigotimes_2 P(Y_{n+1},\ldots,Y_N,X_{n+1},\ldots,X_N)\\ &=\frac{P(X_{n+1}|X_n)}{P(X_{n+1})}\cdot P(Y_{n+1},\ldots,Y_N,X_{n+1}).\end{aligned}$$

This expression is also independent of the summation variable in $\bigotimes_1$. Thus

$$S_1\cdot S_2 = \frac{P(X_{n+1}|X_n)\,P\left(Y_{n+1}^N,X_{n+1}\right)}{P(X_{n+1})}\bigotimes_1 P\left(X^{(n)}\right)P\left(Y_0,\ldots,Y_n|X^{(n)}\right).$$

Obviously, by the definition of $\bigotimes_1$,

$$\begin{aligned}\bigotimes_1 P\left(X^{(n)}\right)\cdot P\left(Y_0,\ldots,Y_n|X^{(n)}\right) &= \bigotimes_1 P(Y_0,\ldots,Y_n,X_0,X_1,\ldots,X_n)\\ &= P(Y_0,\ldots,Y_n,X_n).\end{aligned}$$

In summary we have obtained the left hand side of (5.7) as

$$\frac{1}{P(X_n,X_{n+1})}\cdot P(Y_0,\ldots,Y_n,X_n)\cdot\frac{P(X_{n+1}|X_n)}{P(X_{n+1})}\cdot P(Y_{n+1},\ldots,Y_N,X_{n+1}).$$

Since

$$\frac{1}{P(X_n,X_{n+1})}\cdot\frac{P(X_{n+1}|X_n)}{P(X_{n+1})} = \frac{1}{P(X_n)\cdot P(X_{n+1})}$$

we obtain the left hand side of (5.7) as

$$\frac{P(Y_0,\ldots,Y_n,X_n)}{P(X_n)}\cdot\frac{P(Y_{n+1},\ldots,Y_N,X_{n+1})}{P(X_{n+1})},$$

which is equal to the right hand side of (5.7). ■

13.6 Final Time Sufficiency

The next proposition, also found in (MacDonald and Zucchini 1997), will be invoked later in finding the recursion for the backward variables in both the Baum–Petrie formulation and in Devijver's rescaling found in the next chapter.

Proposition 13.6.1 *For all integers n and m such that $0 \le n \le m \le N$*

$$P(Y_m,\ldots,Y_N \mid X_n,\ldots,X_m) = P(Y_m,\ldots,Y_N \mid X_m). \qquad (6.1)$$

Proof: We start again by writing the left hand side of (5.7) as

$$\frac{1}{P(X_n,\ldots,X_m)}\bigotimes_1\bigotimes_2 P\left(Y_m,\ldots,Y_N|X^{(N)}\right)\cdot P\left(X^{(N)}\right)$$

where $\bigotimes_1$ designates summation over $j_0,\ldots,j_{n-1}$, and $\bigotimes_2$ designates summation over $j_{m+1},\ldots,j_N$. Proposition 13.3.1 gives by (3.2) ($n=0$)

$$P\left(Y_m,\ldots,Y_N|X^{(N)}\right) = P(Y_m,\ldots,Y_N|X_m,\ldots,X_N).$$

This is independent of the summation variable in $\bigotimes_1$. We evaluate first the summation $\bigotimes_1$, obtaining

$$\begin{aligned}\frac{1}{P(X_n,\ldots,X_m)}\cdot\bigotimes_1 P\left(X^{(N)}\right) &= \frac{P(X_n,\ldots,X_N)}{P(X_n,\ldots,X_m)}\\ &= P(X_{m+1},\ldots,X_N|X_n,\ldots,X_m).\end{aligned}$$

Therefore the overall sum to be computed is

$$\bigotimes_2 P(Y_m,\ldots,Y_N|X_m,\ldots,X_N)\cdot P(X_{m+1},\ldots,X_N|X_n,\ldots,X_m).$$

We use the Markov property once more to obtain

$$P(X_{m+1},\dots,X_N|X_n,\dots,X_m) = P(X_{m+1},\dots,X_N|X_m),$$

so that

$$\begin{aligned}&\bigotimes_2 P(Y_m,\dots,Y_N|X_m,\dots,X_N)\cdot P(X_{m+1},\dots,X_N|X_n,\dots,X_m)\\ &= \bigotimes_2 \frac{P(Y_m,\dots,Y_N,X_m,\dots,X_N)}{P(X_m,\dots,X_N)}\frac{P(X_m,\dots,X_N)}{P(X_m)}\\ &= \frac{1}{P(X_m)}\bigotimes_2 P(Y_m,\dots,Y_N,X_m,\dots,X_N).\end{aligned}$$

Since $\bigotimes_2$ is a summation over the values of $X_{m+1},\dots,X_N$, we obtain

$$\frac{1}{P(X_m)}\bigotimes_2 P(Y_m,\dots,Y_N,X_m,\dots,X_N) = \frac{1}{P(X_m)}P(Y_m,\dots,Y_N,X_m),$$

which is the assertion as claimed. ∎

13.7 Proof of Derin's Backwards Recursion

We finally prove Proposition 13.2.1. In order to simplify notation we write the conditional probability in (2.1) as

$$\widehat{\pi}_j(n|N) = P\left(X_n = j|Y^{(N)}\right).$$

By another marginalization and an extended version of the chain rule (2.16) in Chapter 1 we obtain

$$\begin{aligned}\widehat{\pi}_j(n|N) &= \sum_{k=1}^{J} P\left(X_n = j, X_{n+1} = k|Y^{(N)}\right)\\ &= \sum_{k=1}^{J} P\left(X_n = j|X_{n+1} = k, Y^{(N)}\right) P\left(X_{n+1} = k|Y^{(N)}\right)\\ &= \sum_{k=1}^{J} P\left(X_n = j|X_{n+1} = k, Y^{(N)}\right)\widehat{\pi}_k(n+1|N)\\ &= \sum_{k=1}^{J} \frac{P\left(X_n = j, X_{n+1} = k, Y^{(N)}\right)}{P\left(X_{n+1} = k, Y^{(N)}\right)}\widehat{\pi}_k(n+1|N). \qquad (7.2)\end{aligned}$$

In the numerator inside the summation in (7.2) we have

$$\begin{aligned}&P\left(X_n = j, X_{n+1} = k, Y^{(N)}\right)\\ &= P\left(Y^{(N)}|X_n = j, X_{n+1} = k\right) P(X_n = j, X_{n+1} = k)\\ &= P\left(Y^{(n)}|X_n = j\right) P(Y_{n+1},\dots,Y_N|X_{n+1} = k)\cdot P(X_n = j)\cdot a_{j|k}\end{aligned}$$

using the factorization (5.1) in Proposition 13.5.1 and the definition of $a_{j|k}$. Then, since $P\left(Y^{(n)}|X_n = j\right)P(X_n = j) = P\left(X_n = j|Y^{(n)}\right)P\left(Y^{(n)}\right)$, we have obtained

$$\begin{aligned}&P\left(X_n = j, X_{n+1} = k, Y^{(N)}\right)\\ &= P\left(X_n = j|Y^{(n)}\right) P\left(Y^{(n)}\right) P(Y_{n+1},\dots,Y_N|X_{n+1} = k)\cdot a_{j|k}. \qquad (7.3)\end{aligned}$$

For the denominator inside the summation in (7.2) it holds that

$$\begin{aligned}P\left(X_{n+1} = k, Y^{(N)}\right) &= P\left(Y^{(N)}|X_{n+1} = k\right) P(X_{n+1} = k)\\ &= P\left(Y^{(n)}|X_{n+1} = k\right) P(Y_{n+1},\dots,Y_N|X_{n+1} = k) P(X_{n+1} = k), \qquad (7.4)\end{aligned}$$

where we have invoked (5.6) in Proposition 13.5.3. Next, from the identity

$$P\left(Y^{(n)}|X_{n+1} = k\right) P(X_{n+1} = k) = P\left(X_{n+1} = k|Y^{(n)}\right) P\left(Y^{(n)}\right)$$

we obtain

$$\begin{aligned}&P\left(X_{n+1} = k, Y^{(N)}\right)\\ &= P\left(X_{n+1} = k|Y^{(n)}\right) P\left(Y^{(n)}\right) P(Y_{n+1},\dots,Y_N|X_{n+1} = k). \qquad (7.5)\end{aligned}$$

Therefore we obtain in from (7.2)–(7.5) and (2.1) that

$$\begin{aligned}\widehat{\pi}_j(n|N) &= \sum_{k=1}^{J} \frac{P\left(X_n = j|Y^{(n)}\right)\cdot a_{j|k}}{P\left(X_{n+1} = k|Y^{(n)}\right)}\widehat{\pi}_k(n+1|N)\\ &= \widehat{\pi}_j(n|n)\cdot\sum_{k=1}^{J}\frac{a_{j|k}}{\widehat{\pi}_k(n+1|n)}\widehat{\pi}_k(n+1|N),\end{aligned}$$

which is the result claimed in (2.2). ∎