The Backward Algorithm

Observation: $p(o_1 \ldots o_T \mid \lambda) = \sum_{i=1}^{N} \sum_{S_2 \ldots S_T} p(o_1 \ldots o_T, S_1 = s_i, S_2 \ldots S_T)$

$$= \sum_{i=1}^{N} \pi_i b_i(o_1) \sum_{S_2 \ldots S_T} p(o_2 \ldots o_T, S_2 \ldots S_T \mid S_1 = s_i)$$

($o_1 \ldots o_t$ already generated)

Algorithm:

$$\boxed{\beta_t(i) = \sum_{S_{t+1} \ldots S_T} p(o_{t+1} \ldots o_T, S_{t+1} \ldots S_T \mid S_t = s_i)}$$

← **Starting from state s_i**
Generating $o_{t+1} \ldots o_T$

$$= \sum_{S_{t+1} \ldots S_T} p(o_{t+2} \ldots o_T, S_{t+2} \ldots S_T \mid S_{t+1}) p(S_{t+1} \mid S_t = s_i) p(o_{t+1} \mid S_{t+1})$$

$$= \sum_{j=1}^{N} a_{ij} b_j(o_{t+1}) \sum_{S_{t+2} \ldots S_T} p(o_{t+2} \ldots o_T, S_{t+2} \ldots S_T \mid S_{t+1} = s_j)$$

$$= \sum_{j=1}^{N} a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)$$

Complexity: $O(TN^2)$

The data likelihood is

$$p(o_1 \ldots o_T \mid \lambda) = \sum_{i=1}^{N} \pi_i b_i(o_1) \beta_1(i) = \sum_{i=1}^{N} \alpha_1(i) \beta_1(i) = \sum_{i=1}^{N} \alpha_t(i) \beta_t(i) \quad for\ any\ t$$