Markov chain formulas:

Basic rules of Markov chains:

- $P(X_t \mid X_{t-1}, X_{t-2}, X_{t-3}, \ldots) = P(X_t \mid X_{t-1})$ *(what happens at time t only depends on $t-1$)*
- $P(X_t \mid X_{t-1}) = P(X_1 \mid X_0)$ *(transition probability distributions are all identical)*

Probability of a sequence of states from beginning: $P(x_{0:t}) = P(x_0) \prod_{i=1}^{t} P(x_i \mid x_{i-1})$

Probability of the next state (Mini-forward algorithm): $P(x_{t+1}) = \sum_{x_t} P(x_{t+1} \mid x_t) P(x_t)$

In vector notation: $P(X_{t+1}) = v_{t+1} = v_t \cdot T$

$P(X_{t+n}) = v_{t+n} = v_t \cdot T^n$

where v_t is a row vector containing the probability distribution of the Markov chain being in each state at time t.

$v_0 = P(X_0) =$ initial probability distribution of the Markov chain.

HMM (Hidden Markov model) formulas:

Basic rules of HMMs:

- Same two rules as for Markov chains, plus:
- $P(E_t \mid X_{0:t}, E_{0:t-1}) = P(E_t \mid X_t)$ *(evidence at time t depends only on the state at time t)*
- $P(E_t \mid X_t) = P(E_1 \mid X_1)$ *(evidence probability distributions are all identical)*

Forward algorithm: Probability of the next state given all evidence up to this point from beginning:

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

In vector notation: $P(X_{t+1} \mid e_{1:t}) = f_{1:t+1} = \alpha \cdot f_{1:t} \cdot T \cdot O_{t+1}$

$$P(X_{t+1} \mid e_{1:t}) = f_{t+1} = \alpha \cdot f_t \cdot T \cdot O_{t+1}$$

[Assume evidence begins from timestep 1 if not specified.]

where $f_{t_1:t_2}$ is a row vector containing the probability distribution of the Markov chain being in each state at time t_2, given all the evidence between times t_1 and t_2. If $t_1 = 1$, we can drop it in the notation and just use f_t.

$f_{1:0} = f_0 = P(X_0) =$ initial probability distribution of the HMM.

Smoothing (forward-backward algorithm): Probability of a *past* state given evidence up to the present:

$$\begin{aligned} P(X_k \mid e_{1:t}) &= P(X_k \mid e_{1:k}, e_{k+1:t}) \\ &= \alpha \cdot P(X_k \mid e_{1:k}) P(e_{k+1:t} \mid X_k, e_{1:k}) && \text{(Bayes' rule)} \\ &= \alpha \cdot P(X_k \mid e_{1:k}) P(e_{k+1:t} \mid X_k) && \text{(conditional indep)} \\ &= \alpha \cdot f_{1:k} \times b_{k+1:t} && \text{($\times$ is pointwise multiplication of vectors)} \end{aligned}$$

where the f-vectors are as above and the b-vectors are $b_{k:t} = T \cdot O_k \cdot b_{k+1:t}$ with $b_{t+1:t} = \mathbf{1}$ (a column vector of all 1's).

To use the algorithm to compute $P(X_k \mid e_{1:t})$:

- Compute the forward probabilities (f-vectors) from f_0 up to f_t
- Compute the backward probabilities (b-vectors) from $b_{t+1:t}$ back to $b_{k+1:t}$
- Multiply pairs of f- and b-vectors (using element-by-element multiplication within each vector).