1 Basic Markov Chains

Definition 1. A sequence of random variables $X_1, X_2, ...$ taking values in a state space S is a **Markov chain** if it satisfies the Markov property $\forall t$:

$$\mathbb{P}(X_t = x_t | X_1 = x_1, ..., X_{t-1} = x_{t-1}) = \mathbb{P}(X_t = x_t | X_{t-1} = x_{t-1}).$$

That is, the value of X_t depends **only** on the value of X_{t-1}.

Definition 2. A Markov chain $X_1, X_2, ...$ is **time homogeneous** iff the transition probabilities don't depend on t:

$$\mathbb{P}(X_t = j | X_{t-1} = i) = \mathbb{P}(X_{t'} = j | X_{t'-1} = i) \quad \forall t, t'.$$

2 Hidden Markov Models

Definition 3. A **hidden Markov Model** involves an unobservable Markov chain $X_1, X_2, ...$ and a sequence of observations $Y_1, Y_2, ...$ each from an observation space $O = \{o_1, ..., o_M\}$ such that:

$$\mathbb{P}(Y_t = y_t | X_1 = x_1, ..., X_t = x_t, Y_1 = y_1, ..., Y_{t-1} = y_{t-1}, Y_{t+1} = y_{t+1}, ...) = \mathbb{P}(Y_t = y_t | X_t = x_t).$$

That is, Y_t depends **only** on X_t.

An HHM with hidden state space $S = [N]$ and M possible observations can be parameterised fully as $\lambda = (\pi, A, B)$ where:

- $\pi \in \mathbb{R}^N$ is the **initial distribution** over state space S:
$$\pi_i = \mathbb{P}(X_1 = i).$$
- $A \in \mathbb{R}^{N \times N}$ gives us the **transition probabilities**:
$$a_{ij} = \mathbb{P}(X_t = j | X_{t-1} = i).$$
- $B = \{b_j(o_k) : j \in [N], k \in [M]\}$ gives us the **emission/observation probabilities**:
$$b_j(o_k) = \mathbb{P}(Y_t = o_k | X_t = j).$$

The mini-lecture will focus on **filtering**, **smoothing** and **parameter estimation**. In the discussion of these topics we will be using the following quantities:

- Forward probabilities:
$$\alpha_t(i) = \mathbb{P}(Y_1, ..., Y_t, X_t = i | \lambda)$$
- Backward probabilities:
$$\beta_t(i) = \mathbb{P}(Y_{t+1}, ..., Y_T | X_t = i, \lambda)$$
- Smoothed probabilities
$$\gamma_t(i) = \mathbb{P}(X_t = i | Y, \lambda) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^N \alpha_t(j)\beta_t(j)}$$
- Discussion of parameter estimation via the Baum-Welch algorithm will also require the following probability:
$$\xi_t(i, j) = \mathbb{P}(X_t = i, X_{t+1} = j | Y, \lambda)$$