Tutorial: Markov Chain Monte Carlo (MCMC)

Maximum likelihood estimation for Markov chains

- We use MLE to estimate A from data $\mathcal{D} = \{x^{(1)}, ..., x^{(N)}\}$.
- Likelihood of any particular sentence $x^{(i)}$ of length T_i:

$$p(x^{(i)}|\theta) = \prod_{j=1}^{K} \pi_j^{\mathbb{1}[x_1^{(i)}=j]} \prod_{t=2}^{T_i} \prod_{j=1}^{K} \prod_{k=1}^{K} A_{jk}^{\mathbb{1}[x_t^{(i)}=k, x_{t-1}^{(i)}=j]}$$

Maximum likelihood estimation for Markov chains

- We use MLE to estimate A from data $\mathcal{D} = \{x^{(1)}, ..., x^{(N)}\}$.
- Likelihood of any particular sentence $x^{(i)}$ of length T_i:

$$p(x^{(i)}|\theta) = \prod_{j=1}^{K} \pi_j^{\mathbb{1}[x_{\mathbf{1}}^{(i)}=j]} \prod_{t=2}^{T_i} \prod_{j=1}^{K} \prod_{k=1}^{K} A_{jk}^{\mathbb{1}[x_t^{(i)}=k, x_{t-\mathbf{1}}^{(i)}=j]}$$

- Log-likelihood of $\mathcal{D}$ (all sentences treated as independent):

$$\log p(\mathcal{D}|\theta) = \sum_{i=1}^{N} \log p(x^{(i)}|\theta) = \sum_j N_j^1 \log \pi_j + \sum_j \sum_k N_{jk} \log A_{jk}$$

- where we define the counts:

$$N_j^1 = \sum_{i=1}^{N} \mathbb{1}[x_1^{(i)} = j], \qquad N_{jk} = \sum_{i=1}^{N} \sum_{t=1}^{T_i-1} \mathbb{1}[x_t^{(i)} = j, x_{t+1}^{(i)} = k].$$

Maximum likelihood estimation for Markov chains

- We use MLE to estimate A from data $\mathcal{D} = \{x^{(1)}, ..., x^{(N)}\}$.
- Likelihood of any particular sentence $x^{(i)}$ of length T_i:

$$p(x^{(i)}|\theta) = \prod_{j=1}^{K} \pi_j^{\mathbb{1}[x_1^{(i)}=j]} \prod_{t=2}^{T_i} \prod_{j=1}^{K} \prod_{k=1}^{K} A_{jk}^{\mathbb{1}[x_t^{(i)}=k, x_{t-1}^{(i)}=j]}$$

- Log-likelihood of $\mathcal{D}$ (all sentences treated as independent):

$$\log p(\mathcal{D}|\theta) = \sum_{i=1}^{N} \log p(x^{(i)}|\theta) = \sum_j N_j^1 \log \pi_j + \sum_j \sum_k N_{jk} \log A_{jk}$$

- where we define the counts:

$$N_j^1 = \sum_{i=1}^{N} \mathbb{1}[x_1^{(i)} = j], \qquad N_{jk} = \sum_{i=1}^{N} \sum_{t=1}^{T_i - 1} \mathbb{1}[x_t^{(i)} = j, x_{t+1}^{(i)} = k].$$

- The MLE is given as: $\hat{\pi}_j = \frac{N_j^1}{\sum_j N_j^1}$ $\hat{A}_{jk} = \frac{N_{jk}}{\sum_k N_{jk}}.$