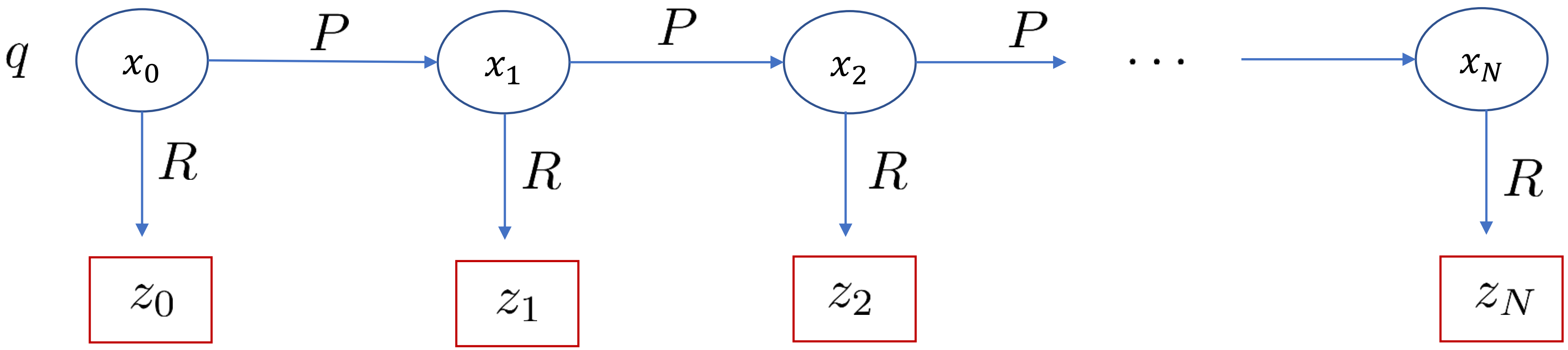
Hidden Markov Models (HMM)

- Consider a Markov Chain with finite number states and transition probabilities:

$$p_{i,j} = \Pr(x_{k+1} = j | x_k = i), \forall k = \{0, ..., N-1\}$$

$$q_i = \Pr(x_0 = i)$$

- A Hidden Markov Model (HMM) can be view schematically as follows:



- Where:

$$r(z|i) = r(z_{k+1} = z, x_k = i)$$

Inference in Hidden Markov Models

- We saw on the previous class how to answer the following questions in the HMM:

1. Computing $\Pr(x_k|z_0, ..., z_k)$ for all possible values of x_k. This is called the *Filtering Problem*.
2. Computing $\Pr(x_k|z_0, ..., z_s), k > s$ for all possible values of x_k. This is called the *Prediction Problem*.
3. Computing $\Pr(x_k|z_0, ..., z_s), k < s$ for all possible values of x_k. This is called the *Smoothing Problem*.

Inference in Hidden Markov Models

- Namely, we developed the Forward-Backward Algorithm:

$$\alpha(x_{k+1}) = \sum_{x_k} \alpha(x_k) p_{x_k, x_{k+1}} r(z_{k+1}|x_{k+1}) \qquad \alpha(x_0) = r(z_0|x_0) q(x_0)$$

$$\beta(x_k) = \sum_{x_{k+1}} \beta(x_{k+1}) p_{x_k, x_{k+1}} r(z_{k+1}|x_{k+1}) \qquad \beta(x_N) = 1$$

- And we saw that we can replace the backward recursion by:

$$\gamma(x_k) = \sum_{x_{k+1}} \frac{\alpha(x_k) p_{x_k, x_{k-1}}}{\sum_{x_k} \alpha(x_k) p_{x_k, x_{k-1}}} \gamma(x_{k+1}) \qquad \gamma(x_N) = \alpha(x_N)$$

Inference in Hidden Markov Models

- And we can make, for example, the following inferences:

$$\Pr(Z_N) = \sum_{x_N} \alpha(x_N)\beta(x_N) = \sum_{x_N} \alpha(x_N)$$

$$\Pr(x_k | Z_N) = \frac{\alpha(x_k)\beta(x_k)}{\Pr(Z_N)}$$

$$\Pr(x_k | z_0, ..., z_k) = \frac{\alpha(x_k)}{\Pr(Z_N)}$$

- Note that all these computations depend on P, R, and q!

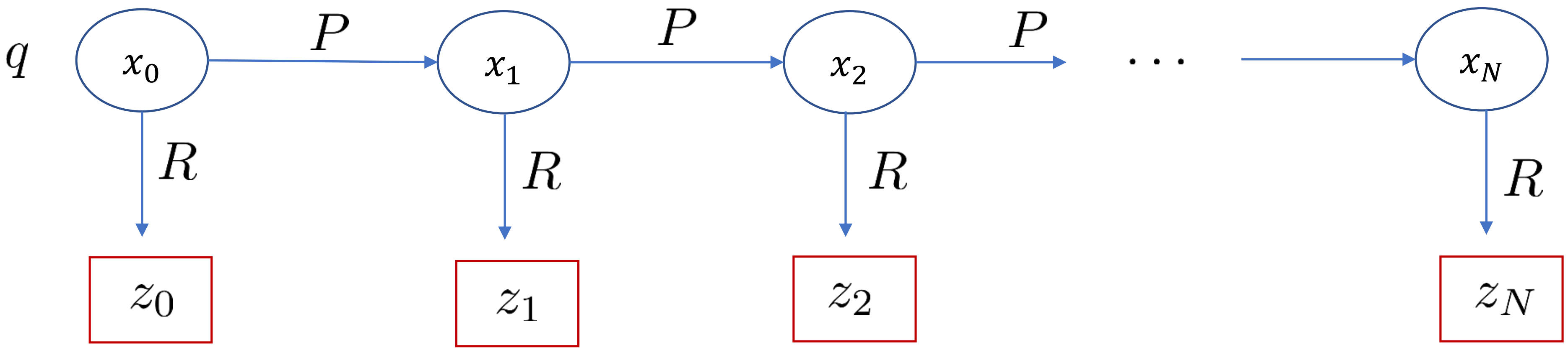
Hidden Markov Models (HMM)

- Consider a Markov Chain with finite number states and transition probabilities:

$$p_{i,j} = \Pr(x_{k+1} = j | x_k = i), \forall k = \{0, ..., N-1\}$$

$$q_i = \Pr(x_0 = i)$$

- A Hidden Markov Model (HMM) can be view schematically as follows:



- Where:

$$r(z|i) = r(z_{k+1} = z, x_k = i)$$

Estimation in Hidden Markov Models

- We will first explore the full information case, where we have a data set X, of hidden state sequences and a data set Z, of observations sequences.

- For simplicity we will focus on the multinomial case:

  - We let the hidden state x_k be a vector with M components, and if the underlying markov chain is in state i at stage k, then we set the i'th element of x_k to be one and all the other elements to be zero.

  - We let the observation z_k be a vector with M components, such that only a single component j is equal to one and all the other elements are equal to zero. We let $r_{i,j}$ to denote the conditional probability that the j'th component of z_k is equal to one, **given** that the I'th component of x_k is one, that is:

$$r_{i,j} = \Pr(z_k^{(j)} = 1 | x_k^{(i)} = 1)$$

Estimation in Hidden Markov Models

- Our goal to estimate the HMM parameters $\theta = (P, R, q)$.

- For that we will perform MLE, by first writing the joint probability of observing our data set (X, Z):

$$\Pr(X, Z|\theta) = \prod_{l=1}^{L} \Pr(x_0^l, ..., x_N^l, z_0^l, ..., z_N^l|\theta)$$

- And using the Markov Property:

$$\Pr(X, Z|\theta) = \prod_{l=1}^{L} q(x_0^l) \prod_{k=1}^{N-1} (p_{x_k^l, x_{k+1}^l}) \prod_{k=1}^{N-1} (r(z_k^l|x_k^l)$$

Estimation in Hidden Markov Models

- Now we introduce the idea of counts: We will count the transitions that we see in our data set (X, Z):

$$m_{i,j} = \sum_{l=1}^{L} \sum_{k=0}^{N-1} x_k^{l,i} x_{k+1}^{l,j}$$

$$\eta_{i,j} = \sum_{l=1}^{L} \sum_{k=0}^{N-1} x_k^{l,i} z_k^{l,j}$$

$$v_i = \sum_{l=1}^{L} x_0^{l,i}$$

Estimation in Hidden Markov Models

- Now we can substitute in the likelihood equation:

$$\Pr(X, Z|\theta) = \prod_{l=1}^{L} \prod_{i=1}^{p} [q_i]^{x_0^{l,i}} \prod_{k=1}^{N-1} \prod_{i=1}^{M} \prod_{j=1}^{M} [p_{i,j}]^{x_k^{l,i} x_{k+1}^{l,j}} \prod_{k=1}^{N-1} \prod_{i=1}^{M} \prod_{j=1}^{M} [r_{i,j}]^{x_k^{l,i} z_k^{l,j}}$$

- Taking the log:

$$\ln(\Pr(X, Z|\theta)) = \sum_{l=1}^{L} \left(\sum_{i=1}^{p} x_0^{l,i} \ln(q_i) + \sum_{k=1}^{N-1} \sum_{i=1}^{M} \sum_{j=1}^{M} x_k^{l,i} x_{k+1}^{l,j} \ln(p_{i,j}) + \sum_{k=1}^{N-1} \sum_{i=1}^{M} \sum_{j=1}^{M} x_k^{l,i} z_k^{l,j} \ln(r_{i,j}) \right)$$

- Rearranging and using our counts:

$$\ln(\Pr(X, Z|\theta)) = \sum_{i=1}^{p} v_i \ln(q_i) + \sum_{i=1}^{M} \sum_{j=1}^{M} m_{i,j} \ln(p_{i,j}) + \sum_{i=1}^{M} \sum_{j=1}^{M} \eta_{i,j} \ln(r_{i,j})$$

Estimation in Hidden Markov Models

- The counts m, η, v are the **sufficient statistics** for our estimation problem, that is, they describe all the information necessary to carry out the estimation. We can then write the optimization problem:

$$
\begin{aligned}
\max_{P,R,q} \sum_{i=1}^{p} v_i \ln(q_i) &+ \sum_{i=1}^{M}\sum_{j=1}^{M} m_{i,j}\ln(p_{i,j}) + \sum_{i=1}^{M}\sum_{j=1}^{M} \eta_{i,j}\ln(r_{i,j}) \\
\text{s.t.:} \quad & \sum_{i=1}^{M} q_i = 1 \\
& \sum_{j=1}^{M} p_{i,j} = 1 \\
& \sum_{j=1}^{M} r_{i,j} = 1 \\
& p_{i,j} \geq 0, r_{i,j} \geq 0, q_i \geq 0, \forall i, j \in \{1, ..., M\}
\end{aligned}
$$

Estimation in Hidden Markov Models

- Using the Lagrange Multiplies Method, one can verify that the optimal solution is given by:

$$p^*_{i,j} = \frac{m_{i,j}}{\sum_{k=1}^{M} m_{i,k}}$$

$$r^*_{i,j} = \frac{\eta_{i,j}}{\sum_{k=1}^{M} \eta_{i,k}}$$

$$q^*_{i} = \frac{v_{i}}{\sum_{k=1}^{M} v_{k}}$$

- Note that the quantities are nonnegative and sum to one.