

732A96 Advanced Machine Learning  
Graphical Models and Hidden Markov Models

Jose M. Peña  
IDA, Linköping University, Sweden

Lecture 6: Autoregressive and Explicit-Duration Hidden Markov Models

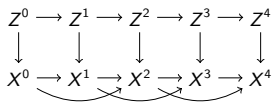
# Contents

- Autoregressive Hidden Markov Models
  - Definition
  - Learning
  - Forward-Backward Algorithm
- Explicit-Duration Hidden Markov Models
  - Definition

- ▶ Main source
  - ▶ Chiappa, S. Explicit-Duration Markov Switching Models. *Foundations and Trends in Machine Learning* 7, 803-886, 2014. Sections 1-3.3.
- ▶ Additional source
  - ▶ Bishop, C. M. *Pattern Recognition and Machine Learning*. Springer, 2006. Chapter 13.1-13.2.

# Autoregressive Hidden Markov Models: Definition

- ▶ To overcome the poor modeling of long range correlations in HMMs, by allowing  $pa_G(X^t) \neq \emptyset$ .



- ▶ Hereinafter, we focus on the simplest AR-HMM, i.e.  $pa_G(X^t) = \{X^{t-1}\}$ .

# Autoregressive Hidden Markov Models: Learning

- ▶ Almost the same as for HMMs.
- ▶ Recall that maximizing the log likelihood function over  $x^{0:T}$  is inefficient (no closed form solution) and ineffective (multimodal).
- ▶ Consider instead maximizing its expectation

$$\begin{aligned} E_{Z^{0:T}} [\log p(Z^{0:T}, x^{0:T})] &= \sum_{z^{0:T}} p(z^{0:T} | x^{0:T}) \log p(z^{0:T}, x^{0:T}) \\ &= \sum_{z^{0:T}} p(z^{0:T} | x^{0:T}) [\log \theta_{z^0} + \sum_{t=1}^T \log \theta_{z^{t+1}|z^t} + \sum_{t=1}^T \log \theta_{x^t|z^t, x^{t-1}}] \\ &= \sum_{z^0} p(z^0 | x^{0:T}) \log \theta_{z^0} + \sum_{t=1}^T \sum_{z^t} \sum_{z^{t+1}} p(z^t, z^{t+1} | x^{0:T}) \log \theta_{z^{t+1}|z^t} + \sum_{t=1}^T \sum_{z_t} p(z^t | x^{0:T}) \log \theta_{x^t|z^t, x^{t-1}} \end{aligned}$$

- ▶ Then

- ▶  $\theta_{z^0}^{ML} = \frac{p(z^0 | x^{0:T})}{\sum_{z^0} p(z^0 | x^{0:T})}$
- ▶  $\theta_{z^{t+1}|z^t}^{ML} = \frac{\sum_{t=1}^T p(z^t, z^{t+1} | x^{0:T})}{\sum_{t=1}^T \sum_{z^{t+1}} p(z^t, z^{t+1} | x^{0:T})}$
- ▶  $\theta_{x^t|z^t, x^{t-1}}^{ML} = \frac{\sum_{t=1}^T p(z^t | x^{0:T}) 1_{\{x^t, x^{t-1} \in x^{0:T}\}}}{\sum_{t=1}^T p(z^t | x^{0:T}) 1_{\{x^{t-1} \in x^{0:T}\}}}$

- ▶ Note that computing  $p(Z^0 | x^{0:T})$ ,  $p(Z^t, Z^{t+1} | x^{0:T})$  and  $p(Z^t | x^{0:T})$  requires inference: Forward-backward algorithm.

# Autoregressive Hidden Markov Models: Forward-Backward Algorithm

$$\begin{aligned}
 p(Z^t | x^{0:T}) &= \frac{p(x^{0:t-1}, x^{t+1:T} | Z^t, x^t) p(Z^t, x^t)}{p(x^{0:T})} \\
 &= \frac{p(x^{0:t-1} | Z^t, x^t) p(Z^t, x^t) p(x^{t+1:T} | Z^t, x^t)}{p(x^{0:T})} \text{ by } x^{0:t-1} \perp_p x^{t+1:T} | Z^t \cup x^t \\
 &= \frac{p(x^{0:t}, Z^t) p(x^{t+1:T} | Z^t, x^t)}{p(x^{0:T})} = \frac{\alpha(Z^t) \beta(Z^t)}{\sum_{z^t} \alpha(z^t) \beta(z^t)}
 \end{aligned}$$

$$\begin{aligned}
 p(Z^t, Z^{t+1} | x^{0:T}) &= \frac{p(x^{0:t-1}, x^{t+2:T} | Z^t, Z^{t+1}, x^t, x^{t+1}) p(Z^t, Z^{t+1}, x^t, x^{t+1})}{p(x^{0:T})} \\
 &= \frac{p(x^{0:t-1} | Z^t, x^t) p(x^{t+2:T} | Z^{t+1}, x^{t+1}) p(x^{t+1} | Z^{t+1}, x^t) p(Z^{t+1} | Z^t) p(Z^t, x^t)}{p(x^{0:T})} \\
 &\quad \text{by } x^{0:t-1} \perp_p x^{t+2:T} | Z^t \cup Z^{t+1} \cup x^t \cup x^{t+1} \\
 &\quad \quad x^{0:t-1} \perp_p Z^{t+1} \cup x^{t+1} | Z^t \cup x^t \\
 &\quad \quad x^{t+2:T} \perp_p Z^t \cup x^t | Z^{t+1} \cup x^{t+1} \\
 &\quad \quad x^{t+1} \perp_p Z^t | Z^{t+1} \cup x^t \\
 &\quad \quad Z^{t+1} \perp_p x^t | Z^t \\
 &= \frac{\alpha(Z^t) \beta(Z^{t+1}) p(x^{t+1} | Z^{t+1}, x^t) p(Z^{t+1} | Z^t)}{\sum_{z^t} \sum_{z^{t+1}} \alpha(z^t) \beta(z^{t+1}) p(x^{t+1} | z^{t+1}, x^t) p(z^{t+1} | z^t)}
 \end{aligned}$$

# Autoregressive Hidden Markov Models: Forward-Backward Algorithm

$$\begin{aligned}
 \alpha(\mathbf{Z}^t) &= p(x^t | Z^t, x^{t-1}) p(Z^t, x^{t-1}) p(x^{0:t-2} | Z^t, x^{t-1}) \text{ by } X^{0:t-2} \perp_p X^t | Z^t \cup X^{t-1} \\
 &= p(x^t | Z^t, x^{t-1}) p(x^{0:t-1}, Z^t) = p(x^t | Z^t, x^{t-1}) \sum_{z^{t-1}} p(x^{0:t-1}, Z^t | z^{t-1}) p(z^{t-1}) \\
 &= p(x^t | Z^t, x^{t-1}) \sum_{z^{t-1}} p(x^{0:t-1} | z^{t-1}) p(Z^t | z^{t-1}) p(z^{t-1}) \text{ by } X^{0:t-1} \perp_p Z^t | Z^{t-1} \\
 &= p(x^t | Z^t, x^{t-1}) \sum_{z^{t-1}} p(x^{0:t-1}, z^{t-1}) p(Z^t | z^{t-1}) = p(x^t | Z^t, x^{t-1}) \sum_{z^{t-1}} \alpha(\mathbf{z}^{t-1}) p(Z^t | z^{t-1})
 \end{aligned}$$

$$\alpha(Z^0) = p(x^0 | Z^0) p(Z^0)$$

$$\beta(\mathbf{Z}^t) =$$

$$\begin{aligned}
 &= \sum_{z^{t+1}} p(x^{t+1:T} | z^{t+1}, x^t) p(z^{t+1} | Z^t) \text{ by } X^{t+1:T} \perp_p Z^t | Z^{t+1} \cup X^t \text{ and } X^t \perp_p Z^{t+1} | Z^t \\
 &= \sum_{z^{t+1}} p(x^{t+2:T} | z^{t+1}, x^{t+1}) p(x^{t+1} | z^{t+1}, x^t) p(z^{t+1} | Z^t) \text{ by } X^{t+2:T} \perp_p X^t | Z^{t+1} \cup X^{t+1} \\
 &= \sum_{z^{t+1}} \beta(\mathbf{z}^{t+1}) p(x^{t+1} | z^{t+1}, x^t) p(z^{t+1} | Z^t)
 \end{aligned}$$

$$\beta(Z^T) = 1 \text{ by } p(Z^T | x^{0:T}) = \frac{\alpha(Z^T) \beta(Z^T)}{p(x^{0:T})} = p(Z^T | x^{0:T}) \beta(Z^T)$$

► **Exercise.** Derive the Viterbi algorithm for AR-HMMs.

## Explicit-Duration Hidden Markov Models: Definition