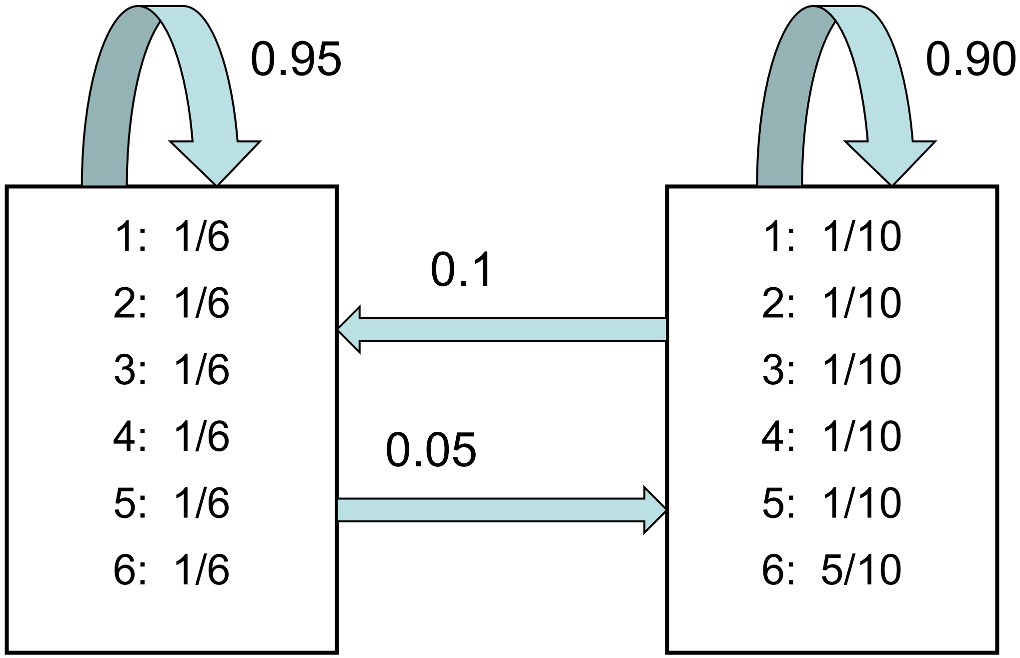
Royal Institute of Technology

DGM APPLICATION HMM

EMISSION & TRANSITION DISTRIBUTIONS



THE JOINT DISTRIBUTION

$$
\begin{aligned}
p(\boldsymbol{x}_{1:T}, \boldsymbol{z}_{1:T+1}) \\
&= p(\boldsymbol{x}_{1:T} | \boldsymbol{z}_{1:T}) p(\boldsymbol{z}_{1:T+1}) \\
&= p(\boldsymbol{z}_1) \left(\prod_{t=1}^{T} p(\boldsymbol{z}_{t+1} | \boldsymbol{z}_t) \right) \left(\prod_{t=1}^{T} p(\boldsymbol{x}_t | \boldsymbol{z}_t) \right)
\end{aligned}
$$

Categorial (or Gaussian)

- ★ Starts in the state $z_1 = k^*$
- ★ When in state z_t
  - emits $p(x_t|z_t)$ $\quad B_{x_t, z_t}$
  - transits to $p(z_{t+1}|z_t)$ $\quad A_{z_{t+1}, z_t}$
- ★ Stops when reaching a stop state
- ★ z_{T+1} is assumed to be stop

The parameters

Now given!!

AN HMM CAN BE SEEN AS A DGM

$$
\begin{tikzcd}
Z_1 \arrow[r] \arrow[d] & Z_2 \arrow[r] \arrow[d] & Z_3 \arrow[r] \arrow[d] & \cdots \arrow[r] & Z_T \arrow[r] \arrow[d] & Z_{T+1} \\
x_1 & x_2 & x_3 & & x_T &
\end{tikzcd}
$$

- Z_i hidden
- X_i observable
- Hidden often not observable when training, never when applying

All the same

State	1	2	3	4
1	A_{11}	A_{21}	A_{31}	A_{41}
2	A_{12}	A_{22}	A_{32}	A_{42}
3	A_{13}	A_{23}	A_{33}	A_{43}
4	A_{14}	A_{24}	A_{34}	A_{44}

$$
\begin{array}{ccccccccccc}
Z_1 & \rightarrow & Z_2 & \rightarrow & Z_3 & \rightarrow & \cdots & \rightarrow & Z_T & \rightarrow & Z_{T+1} \\
\downarrow & & \downarrow & & \downarrow & & & & \downarrow & & \\
x_1 & & x_2 & & x_3 & & & & x_T & &
\end{array}
$$

TRANSITION PROBABILITIES FOR 4 STATES HMM

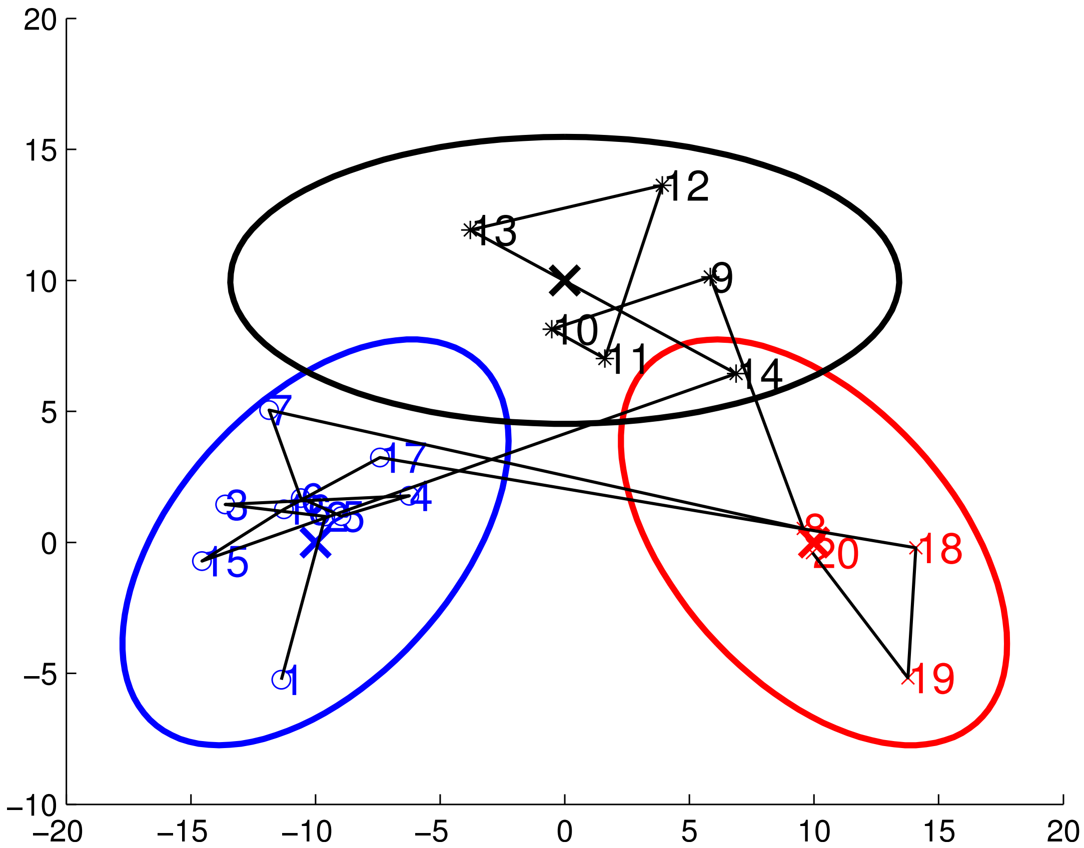
EMISSION PROBABILITIES - HMM WITH 4 STATES & 3 SYMBOLS

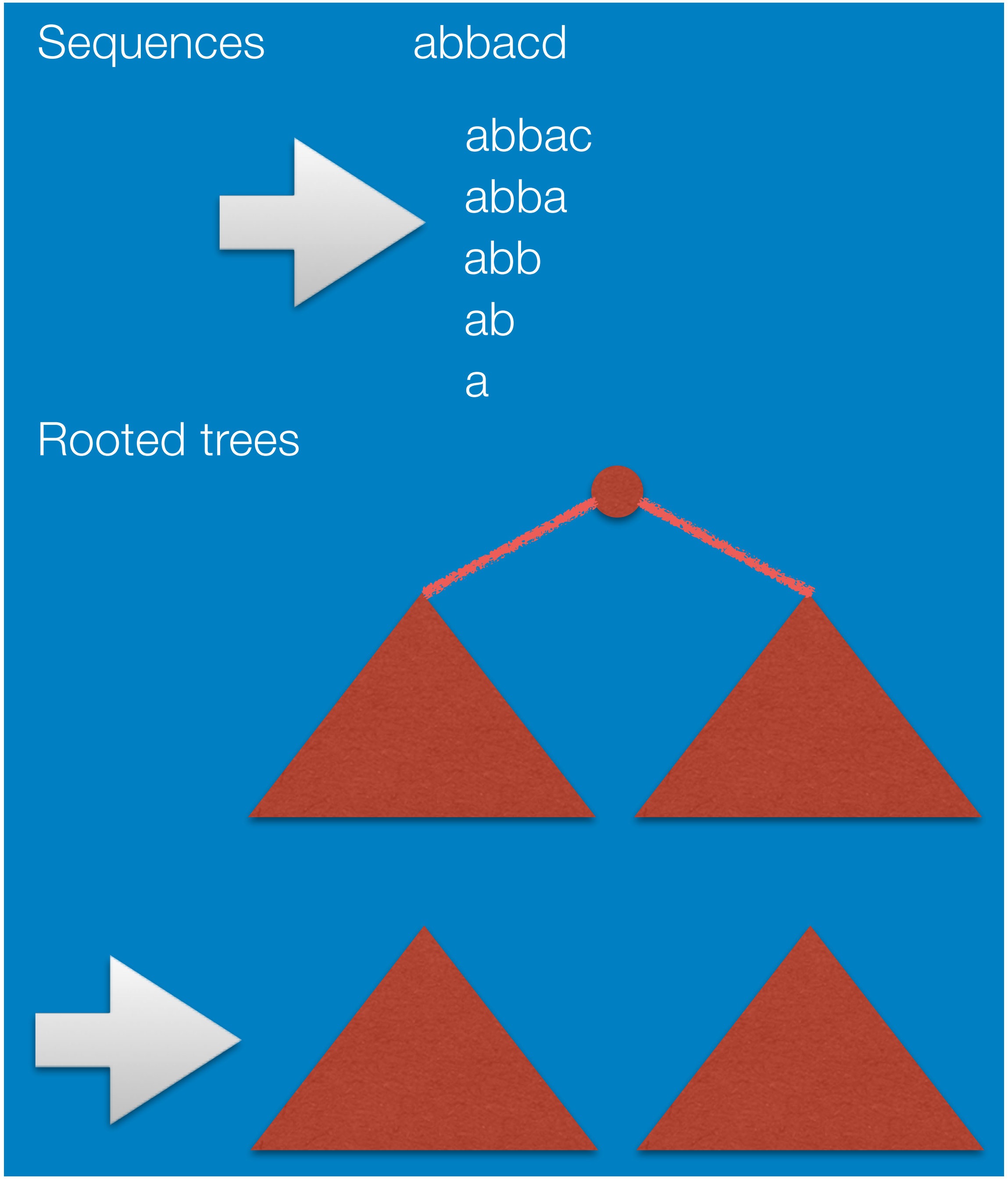
$$
\begin{tikzcd}
Z_1 \arrow[r] \arrow[d] & Z_2 \arrow[r] \arrow[d] & Z_3 \arrow[r] \arrow[d] & \cdots \arrow[r] & Z_T \arrow[r] \arrow[d] & Z_{T+1} \\
x_1 & x_2 & x_3 & & x_T &
\end{tikzcd}
$$

All the same

State\Symb	1	2	3
1	B_{11}	B_{21}	B_{31}
2	B_{12}	B_{22}	B_{32}
3	B_{13}	B_{23}	B_{33}
4	B_{14}	B_{24}	B_{34}

GAUSSIAN EMISSIONS AND HIDDEN STATES





DP

- What is a subproblem?
- What is a subsolution?
- How do we decompose into smaller subproblems?
- How do we combine subsolutions into larger?
- How do we enumerate?
- How many and what time?

DP

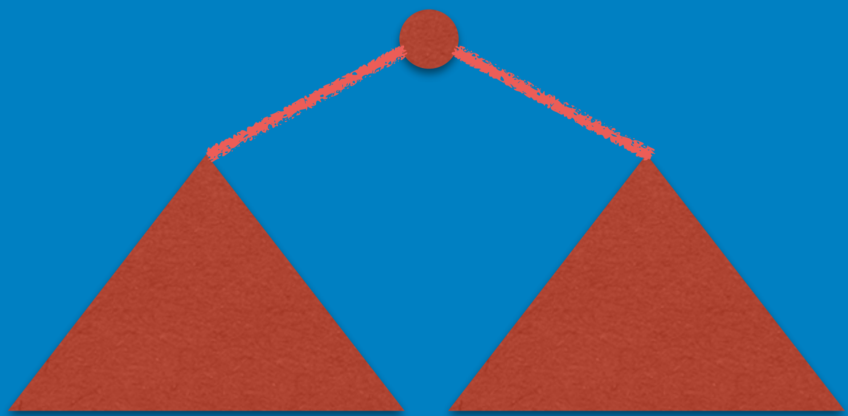
Polynomial many	What is a subproblem?
	What is a subsolution?
Polynomial time	How do we decompose into smaller subproblems?
Polynomial time	How do we combine subsolutions into larger?
Polynomial time	How do we enumerate?
Polynomial time overall	How many and what time?

DP, SUM RULE, & CONDITIONING

Sequences

$$\begin{array}{ccccc} Z_1 & \rightarrow & Z_2 & \rightarrow & Z_3 \\ \downarrow & & \downarrow & & \downarrow \\ x_1 & & x_2 & & x_3 \end{array}$$

Rooted trees



$$P(A) = \sum_i P(A, B = i) = \sum_i P(A \,|\, B = i) P(B = i)$$

- How do we decompose into smaller subproblems?

Dynamic programming is similar to divide + conquer in that it solves a problem by dividing it into sub-problems. However, in the dynamic programming paradigm, the larger problem is solved by solving and remembering overlapping sub-problems, which are reused repeatedly in the process.

Applying sum rule

Notice, by the sum rule,

$$f_t(k) = p(x_{1:t-1}, Z_t = k) = \sum_{k' \in [K]} p(x_{1:t-1}, Z_{t-1} = k', Z_t = k)$$

The set of states

each term in the sum is a probability of an event

$$\begin{array}{ccccccccc}
? & \rightarrow & ? & \rightarrow & ? & \cdots Z_{t-1} = k' & \rightarrow & Z_t = k & \rightarrow ? \cdots \\
\downarrow & & \downarrow & & \downarrow & \downarrow & & \downarrow & \\
x_1 & & x_2 & & x_3 & x_{t-1} & & ? &
\end{array}$$

which, as noted, can be broken into smaller

Forward recursion

$$f_t(k) = \sum_{k'} \underbrace{f_{t-1}(k')}_{\text{smaller}} \underbrace{p(\boldsymbol{x}_{t-1} | \boldsymbol{Z}_{t-1} = k')}_{\text{emission}} \underbrace{p(\boldsymbol{Z}_t = k | \boldsymbol{Z}_{t-1} = k')}_{\text{transition}}$$

Backward variable

Defined by

$$b_t(k) := p(\boldsymbol{x}_{t+1:T} | \boldsymbol{Z}_t = k)$$

“Graphical model”

$$\begin{array}{cccccccc} ? & \rightarrow & Z_t = k & \rightarrow & ? & \rightarrow & \cdots & ? \\ & & \downarrow & & \downarrow & & & \downarrow \\ & & ? & & x_{t+1} & & & x_T \end{array}$$

Backward recursion

$$b_t(k) := p(\boldsymbol{x}_{t+1:T} | \boldsymbol{Z}_t = k)$$

- DP also for the backward variable b_t

$$b_t(k) = \sum_l \underbrace{p(\boldsymbol{Z}_{t+1} = l | \boldsymbol{Z}_t = k)}_{\text{transition}} \underbrace{b_{t+1}(l)}_{\text{"smaller"}} \underbrace{p(\boldsymbol{x}_{t+1} | \boldsymbol{Z}_{t+1} = l)}_{\text{emission}}$$