CSCI x490 Algorithms for Computational Biology

Lecture Note 4 (by Liming Cai)

April 21, 2016

Part IV Probabilistic Models and Learning

Part IV Probabilistic Models and Learning

Probabilistic models

Part IV Probabilistic Models and Learning

Probabilistic models

- Characterization of data observable from a system

Part IV Probabilistic Models and Learning

Probabilistic models

- Characterization of data observable from a system
- Expression of uncertainty of data with probability theory

Part IV Probabilistic Models and Learning

Probabilistic models

- Characterization of data observable from a system
- Expression of uncertainty of data with probability theory
- Automatic learning of the system from data

Part IV Probabilistic Models and Learning

Probabilistic models

- Characterization of data observable from a system
- Expression of uncertainty of data with probability theory
- Automatic learning of the system from data
- Computational inference/prediction of unknown

Part IV Probabilistic Models and Learning

Part IV Probabilistic Models and Learning

1. Parameter Re-estimation for HMMs

Part IV Probabilistic Models and Learning

1. Parameter Re-estimation for HMMs
2. SCFG for Co-evolution

Part IV Probabilistic Models and Learning

1. Parameter Re-estimation for HMMs
2. SCFG for Co-evolution
3. Learning of Markov Networks

Part IV Probabilistic Models and Learning

1. Parameter Re-estimation for HMMs
2. SCFG for Co-evolution
3. Learning of Markov Networks

Part IV Probabilistic Models and Learning

Part IV Probabilistic Models and Learning

1. **Parameter Re-estimation for HMMs**

1. Parameter Re-estimation for HMMs

- HMM (and profile HMM) (introduced)
  topology G: states and transitions
  parameter θ: probability distributions for emissions and transitions

1. Parameter Re-estimation for HMMs

- HMM (and profile HMM) (introduced)
  topology G: states and transitions
  parameter θ: probability distributions for emissions and transitions

- parameter estimation (given G, and data D)

1. Parameter Re-estimation for HMMs

- HMM (and profile HMM) (introduced)
  topology G: states and transitions
  parameter θ: probability distributions for emissions and transitions

- parameter estimation (given G, and data D)

$$\theta^* = \arg\max_{\theta} P(\theta | D)$$

1. Parameter Re-estimation for HMMs

- HMM (and profile HMM) (introduced)
  topology G: states and transitions
  parameter θ: probability distributions for emissions and transitions

- parameter estimation (given G, and data D)

$$\theta^* = \arg\max_{\theta} P(\theta|D)$$

Part IV Probabilistic Models and Learning

1. Parameter Re-estimation for HMMs

- HMM (and profile HMM) (introduced)
  topology G: states and transitions
  parameter θ: probability distributions for emissions and transitions

- parameter estimation (given G, and data D)

$$\theta^* = \arg\max_\theta P(\theta|D)$$

- with structured data (e.g., a multiple alignment from D)
  θ^* can be computed (with prior $P(\theta)$) by Bayes's

Part IV Probabilistic Models and Learning

1. Parameter Re-estimation for HMMs

- HMM (and profile HMM) (introduced)
  topology G: states and transitions
  parameter θ: probability distributions for emissions and transitions

- parameter estimation (given G, and data D)

$$\theta^* = \arg\max_\theta P(\theta|D)$$

- with structured data (e.g., a multiple alignment from D)
  θ^* can be computed (with prior $P(\theta)$) by Bayes's

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

Part IV Probabilistic Models and Learning

1. Parameter Re-estimation for HMMs

- HMM (and profile HMM) (introduced)
  topology G: states and transitions
  parameter θ: probability distributions for emissions and transitions

- parameter estimation (given G, and data D)

$$\theta^* = \arg\max_{\theta} P(\theta|D)$$

- with structured data (e.g., a multiple alignment from D)
  θ^* can be computed (with prior $P(\theta)$) by Bayes's

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} \text{ so } P(\theta|D) \approx P(D|\theta)P(\theta)$$

Part IV Probabilistic Models and Learning

1. Parameter Re-estimation for HMMs

- HMM (and profile HMM) (introduced)
  topology G: states and transitions
  parameter θ: probability distributions for emissions and transitions

- parameter estimation (given G, and data D)

$$\theta^* = \arg\max_{\theta} P(\theta|D)$$

- with structured data (e.g., a multiple alignment from D)
  θ^* can be computed (with prior $P(\theta)$) by Bayes's

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} \text{ so } P(\theta|D) \approx P(D|\theta)P(\theta)$$

with maximum likelihood method

Part IV Probabilistic Models and Learning

1. Parameter Re-estimation for HMMs

- HMM (and profile HMM) (introduced)
  topology G: states and transitions
  parameter θ: probability distributions for emissions and transitions

- parameter estimation (given G, and data D)

$$\theta^* = \arg\max_{\theta} P(\theta|D)$$

- with structured data (e.g., a multiple alignment from D)
  θ^* can be computed (with prior $P(\theta)$) by Bayes's

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} \text{ so } P(\theta|D) \approx P(D|\theta)P(\theta)$$

with maximum likelihood method
  $P(D|\theta)$ is maximized when θ is the frequencies obtained from D.

Part IV Probabilistic Models and Learning

Part IV Probabilistic Models and Learning

- without structural data (e.g., with incomplete data), θ can still be estimated.

- without structural data (e.g., with incomplete data), θ can still be estimated.

  $P(D|\theta_{old})$: probability of data D given the parameter θ_{old}

Part IV Probabilistic Models and Learning

- without structural data (e.g., with incomplete data), θ can still be estimated.

  $P(D|\theta_{old})$: probability of data D given the parameter θ_{old}
  $P(D, e|\theta_{old})$: probability of D with event e given the parameter θ_{old}

Part IV Probabilistic Models and Learning

- without structural data (e.g., with incomplete data), θ can still be estimated.

  $P(D|\theta_{old})$: probability of data D given the parameter θ_{old}
  $P(D, e|\theta_{old})$: probability of D with event e given the parameter θ_{old}

$$P(e|\theta_{new}) = f(e, D, \theta_{old}) = \frac{P(D, e|\theta_{old})}{P(D|\theta_{old})}$$

Part IV Probabilistic Models and Learning

- without structural data (e.g., with incomplete data), θ can still be estimated.

  $P(D|\theta_{old})$: probability of data D given the parameter θ_{old}
  $P(D, e|\theta_{old})$: probability of D with event e given the parameter θ_{old}

$$P(e|\theta_{new}) = f(e, D, \theta_{old}) = \frac{P(D, e|\theta_{old})}{P(D|\theta_{old})}$$

- How to compute $P(D, e|\theta_{old})$ and $P(D|\theta_{old})$?

Part IV Probabilistic Models and Learning

Part IV Probabilistic Models and Learning

- Let sequence $x \in D$, and π represents any path in HMM,

Part IV Probabilistic Models and Learning

- Let sequence $x \in D$, and π represents any path in HMM, then

$$P(x|\theta_{old}) = \sum_{\pi} P(x, \pi|\theta_{old})$$

Part IV Probabilistic Models and Learning

- Let sequence $x \in D$, and π represents any path in HMM, then

$$P(x|\theta_{old}) = \sum_{\pi} P(x, \pi|\theta_{old})$$

  the right-hand-side can be computed with DP similar to one used in the **Viterbi**'s

- Let sequence $x \in D$, and π represents any path in HMM, then

$$P(x|\theta_{old}) = \sum_{\pi} P(x, \pi|\theta_{old})$$

the right-hand-side can be computed with DP similar to one used in the **Viterbi**'s that compute $\max_{\pi} P(x, \pi|\theta_{old})$

Part IV Probabilistic Models and Learning

- Let sequence $x \in D$, and π represents any path in HMM, then

$$P(x|\theta_{old}) = \sum_{\pi} P(x, \pi|\theta_{old})$$

the right-hand-side can be computed with DP similar to one used in the **Viterbi**'s that compute $\max_{\pi} P(x, \pi|\theta_{old})$

$$P(D|\theta_{old}) = \sum_{x \in D} P(x|\theta_{old})$$

Part IV Probabilistic Models and Learning

- Let sequence $x \in D$, and π represents any path in HMM, then

$$P(x|\theta_{old}) = \sum_{\pi} P(x, \pi|\theta_{old})$$

the right-hand-side can be computed with DP similar to one used in the **Viterbi**'s that compute $\max_{\pi} P(x, \pi|\theta_{old})$

$$P(D|\theta_{old}) = \sum_{x \in D} P(x|\theta_{old})$$

The algorithm to compute $\sum_{\pi} P(x, \pi|\theta_{old})$ is **Forward Algorithm**.

Part IV Probabilistic Models and Learning

Part IV Probabilistic Models and Learning

- To compute $P(D, e|\theta_{old})$,

- To compute $P(D, e|\theta_{old})$, note that, for $x \in D$,

$$P(x, e|\theta_{old}) = \sum_{\pi \text{ contains } e} P(x, \pi, e|\theta_{old})$$

Part IV Probabilistic Models and Learning

- To compute $P(D, e|\theta_{old})$, note that, for $x \in D$,

$$P(x, e|\theta_{old}) = \sum_{\pi \text{ contains } e} P(x, \pi, e|\theta_{old})$$

for HMMs, path π containing event e can be expressed as

Part IV Probabilistic Models and Learning

- To compute $P(D, e|\theta_{old})$, note that, for $x \in D$,

$$P(x, e|\theta_{old}) = \sum_{\pi \text{ contains } e} P(x, \pi, e|\theta_{old})$$

for HMMs, path π containing event e can be expressed as

$$\pi = \alpha e \beta$$

where α is a prefix path of π and β is the corresponding suffix path of π.

Part IV Probabilistic Models and Learning

- To compute $P(D, e|\theta_{old})$, note that, for $x \in D$,

$$P(x, e|\theta_{old}) = \sum_{\pi \text{ contains } e} P(x, \pi, e|\theta_{old})$$

for HMMs, path π containing event e can be expressed as

$$\pi = \alpha e \beta$$

where α is a prefix path of π and β is the corresponding suffix path of π. Then

$$P(x, e|\theta_{old}) = \sum_{\alpha e \beta} P(x, \alpha e \beta|\theta_{old})$$

$$P(x, \alpha e \beta|\theta_{old}) = \sum_{j=0}^{n} P(x_{[1..j]}, \alpha|\theta_{old}) P(e|\theta_{old}) P(x_{[j+1..n]}, \beta|\theta_{old})$$

Part IV Probabilistic Models and Learning

- To compute $P(D, e|\theta_{old})$, note that, for $x \in D$,

$$P(x, e|\theta_{old}) = \sum_{\pi \text{ contains } e} P(x, \pi, e|\theta_{old})$$

for HMMs, path π containing event e can be expressed as

$$\pi = \alpha e \beta$$

where α is a prefix path of π and β is the corresponding suffix path of π. Then

$$P(x, e|\theta_{old}) = \sum_{\alpha e \beta} P(x, \alpha e \beta|\theta_{old})$$

$$P(x, \alpha e \beta|\theta_{old}) = \sum_{j=0}^{n} P(x_{[1..j]}, \alpha|\theta_{old}) P(e|\theta_{old}) P(x_{[j+1..n]}, \beta|\theta_{old})$$

where position j partitions the sequence x into two segments $x_{[1..j]}$ and $x_{[j+1..n]}$.

Part IV Probabilistic Models and Learning

To compute

$$P(x, \alpha e \beta | \theta_{old}) = \sum_{j=0}^{n} P(x_{[1..j]}, \alpha | \theta_{old}) P(e | \theta_{old}) P(x_{[j+1..n]}, \beta | \theta_{old})$$

Part IV Probabilistic Models and Learning

To compute

$$P(x, \alpha e \beta | \theta_{old}) = \sum_{j=0}^{n} P(x_{[1..j]}, \alpha | \theta_{old}) P(e | \theta_{old}) P(x_{[j+1..n]}, \beta | \theta_{old})$$

- $P(x_{[1..j]}, \alpha | \theta_{old})$ is already a part of the **Forward Algorithm** to compute $P(x|\theta_{old})$;

To compute

$$P(x, \alpha e \beta | \theta_{old}) = \sum_{j=0}^{n} P(x_{[1..j]}, \alpha | \theta_{old}) P(e | \theta_{old}) P(x_{[j+1..n]}, \beta | \theta_{old})$$

- $P(x_{[1..j]}, \alpha | \theta_{old})$ is already a part of the **Forward Algorithm** to compute $P(x | \theta_{old})$;
- $P(e | \theta_{old})$ is known;

Part IV Probabilistic Models and Learning

To compute

$$P(x, \alpha e \beta | \theta_{old}) = \sum_{j=0}^{n} P(x_{[1..j]}, \alpha | \theta_{old}) P(e | \theta_{old}) P(x_{[j+1..n]}, \beta | \theta_{old})$$

- $P(x_{[1..j]}, \alpha | \theta_{old})$ is already a part of the **Forward Algorithm** to compute $P(x|\theta_{old})$;
- $P(e|\theta_{old})$ is known;
- $P(x_{[j+1..n]}, \beta | \theta_{old})$ is a part of the so-called **Backward Algorithm** that computes $P(x|\theta_{old})$.

Part IV Probabilistic Models and Learning

To compute

$$P(x, \alpha e \beta|\theta_{old}) = \sum_{j=0}^{n} P(x_{[1..j]}, \alpha|\theta_{old}) P(e|\theta_{old}) P(x_{[j+1..n]}, \beta|\theta_{old})$$

- $P(x_{[1..j]}, \alpha|\theta_{old})$ is already a part of the **Forward Algorithm** to compute $P(x|\theta_{old})$;
- $P(e|\theta_{old})$ is known;
- $P(x_{[j+1..n]}, \beta|\theta_{old})$ is a part of the so-called **Backward Algorithm** that computes $P(x|\theta_{old})$.

  Backward Algorithm computes for suffixes instead of prefixes.

Part IV Probabilistic Models and Learning

Examples of event e in a profile HMM (with begin and end states B and E);

Part IV Probabilistic Models and Learning

Examples of event e in a profile HMM (with begin and end states B and E);

- if e is transition $M_i \to M_{i+1}$,

Examples of event e in a profile HMM (with begin and end states B and E);

- if e is transition $M_i \to M_{i+1}$,
  then prefix path α is $B \rightsquigarrow M_i$ and suffix path β is $M_{i+1} \rightsquigarrow E$.

Part IV Probabilistic Models and Learning

Examples of event e in a profile HMM (with begin and end states B and E);

- if e is transition $M_i \to M_{i+1}$,
  then prefix path α is $B \rightsquigarrow M_i$ and suffix path β is $M_{i+1} \rightsquigarrow E$.
- if e is transition $I_i \to D_{i+1}$,

Examples of event e in a profile HMM (with begin and end states B and E);

- if e is transition $M_i \to M_{i+1}$,
  then prefix path α is $B \rightsquigarrow M_i$ and suffix path β is $M_{i+1} \rightsquigarrow E$.

- if e is transition $I_i \to D_{i+1}$,
  then prefix path α is $B \rightsquigarrow I_i$ and suffix path β is $D_{i+1} \rightsquigarrow E$.

Examples of event e in a profile HMM (with begin and end states B and E);

- if e is transition $M_i \to M_{i+1}$,
  then prefix path α is $B \rightsquigarrow M_i$ and suffix path β is $M_{i+1} \rightsquigarrow E$.
- if e is transition $I_i \to D_{i+1}$,
  then prefix path α is $B \rightsquigarrow I_i$ and suffix path β is $D_{i+1} \rightsquigarrow E$.
- if e is emission that M_i emits letter `A`,

Examples of event e in a profile HMM (with begin and end states B and E);

- if e is transition $M_i \to M_{i+1}$,
  then prefix path α is $B \rightsquigarrow M_i$ and suffix path β is $M_{i+1} \rightsquigarrow E$.
- if e is transition $I_i \to D_{i+1}$,
  then prefix path α is $B \rightsquigarrow I_i$ and suffix path β is $D_{i+1} \rightsquigarrow E$.
- if e is emission that M_i emits letter `A`,
  then prefix path α is $B \rightsquigarrow M_i$ and suffix path β is $M_i \rightsquigarrow E$.

Part IV Probabilistic Models and Learning

Examples of event e in a profile HMM (with begin and end states B and E);

- if e is transition $M_i \to M_{i+1}$,
  then prefix path α is $B \rightsquigarrow M_i$ and suffix path β is $M_{i+1} \rightsquigarrow E$.
- if e is transition $I_i \to D_{i+1}$,
  then prefix path α is $B \rightsquigarrow I_i$ and suffix path β is $D_{i+1} \rightsquigarrow E$.
- if e is emission that M_i emits letter `A`,
  then prefix path α is $B \rightsquigarrow M_i$ and suffix path β is $M_i \rightsquigarrow E$.
- etc..

Part IV Probabilistic Models and Learning

Summary of parameter re-estimation for HMMs from given data D

Part IV Probabilistic Models and Learning

Summary of parameter re-estimation for HMMs from given data D

- Assume θ_{old} to be the existing parameter set for an HMM;

Summary of parameter re-estimation for HMMs from given data D

- Assume θ_{old} to be the existing parameter set for an HMM;
- For every event e in the HMM, compute

Part IV Probabilistic Models and Learning

Summary of parameter re-estimation for HMMs from given data D

- Assume θ_{old} to be the existing parameter set for an HMM;
- For every event e in the HMM, compute

$$P(e|\theta_{new}) = \frac{P(D, e|\theta_{old})}{P(D|\theta_{old})}$$

Part IV Probabilistic Models and Learning

Summary of parameter re-estimation for HMMs from given data D

- Assume θ_{old} to be the existing parameter set for an HMM;
- For every event e in the HMM, compute

$$P(e|\theta_{new}) = \frac{P(D, e|\theta_{old})}{P(D|\theta_{old})}$$

where $P(D|\theta_{old}) = \sum_{x \in D} P(x|\theta_{old})$ that can be computed with **Forward**

Part IV Probabilistic Models and Learning

Summary of parameter re-estimation for HMMs from given data D

- Assume θ_{old} to be the existing parameter set for an HMM;
- For every event e in the HMM, compute

$$P(e|\theta_{new}) = \frac{P(D, e|\theta_{old})}{P(D|\theta_{old})}$$

where $P(D|\theta_{old}) = \sum_{x \in D} P(x|\theta_{old})$ that can be computed with **Forward**

and $P(D, e|\theta_{old}) = \sum_{x \in D} P(x, e|\theta_{old})$ that can be computed with both **Forward** and **Backward**.

Summary of parameter re-estimation for HMMs from given data D

- Assume θ_{old} to be the existing parameter set for an HMM;
- For every event e in the HMM, compute

$$P(e|\theta_{new}) = \frac{P(D, e|\theta_{old})}{P(D|\theta_{old})}$$

  where $P(D|\theta_{old}) = \sum_{x \in D} P(x|\theta_{old})$ that can be computed with **Forward**

  and $P(D, e|\theta_{old}) = \sum_{x \in D} P(x, e|\theta_{old})$ that can be computed with both **Forward** and **Backward**.

- iterate the above steps until $\frac{|P(D|\theta_{new}) - P(D|\theta_{old})|}{P(D|\theta_{old})} < \Delta$. for a given constant Δ.

Called Forward-backward algorithm.

Part IV Probabilistic Models and Learning

2. Stochastic context-free grammar

Part IV Probabilistic Models and Learning

2. Stochastic context-free grammar

Part IV Probabilistic Models and Learning

2. **Stochastic context-free grammar**

   - an extension of HMM;

Part IV Probabilistic Models and Learning

2. Stochastic context-free grammar

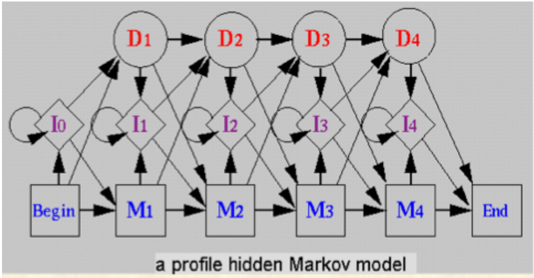
- an extension of HMM;
- with a capability to model correlation and coevolution;

Part IV Probabilistic Models and Learning

2. Stochastic context-free grammar

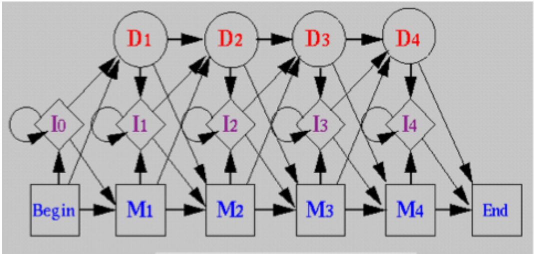
- an extension of HMM;
- with a capability to model correlation and coevolution;
- correlation patterns are limited.

Part IV Probabilistic Models and Learning



We can use a different notation for the HMM.

Part IV Probabilistic Models and Learning

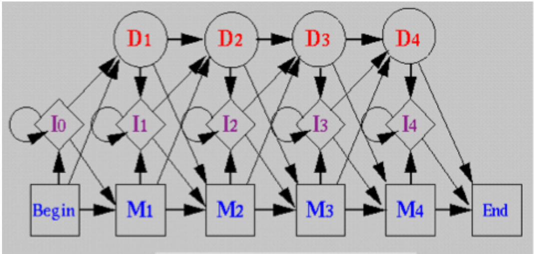


a profile hidden Markov model

We can use a different notation for the HMM.

- $Begin \to eM_1, \quad e \in \{a, c, g, t\}$

Part IV Probabilistic Models and Learning

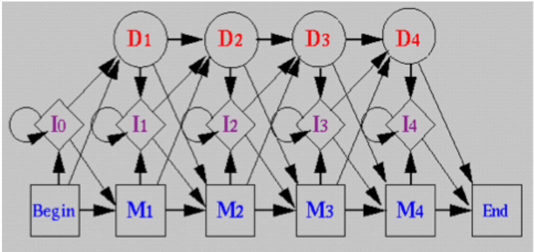


a profile hidden Markov model

We can use a different notation for the HMM.

- $Begin \to eM_1, \quad e \in \{a, c, g, t\}$
  $Begin \to eI_0$

Part IV Probabilistic Models and Learning

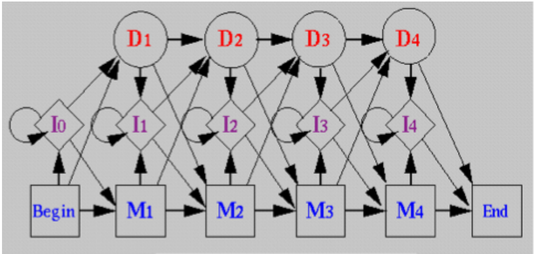


a profile hidden Markov model

We can use a different notation for the HMM.

- $Begin \to eM_1, \quad e \in \{a, c, g, t\}$
  $Begin \to eI_0$
  $Begin \to D_1$

Part IV Probabilistic Models and Learning

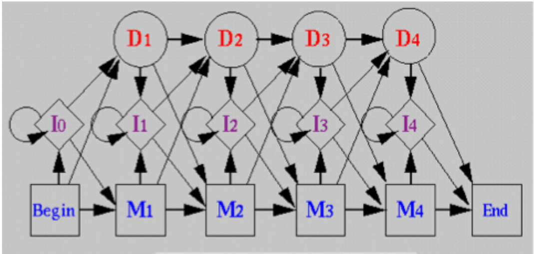


a profile hidden Markov model

We can use a different notation for the HMM.

- $Begin \to eM_1, \quad e \in \{a, c, g, t\}$
  $Begin \to eI_0$
  $Begin \to D_1$

Part IV Probabilistic Models and Learning

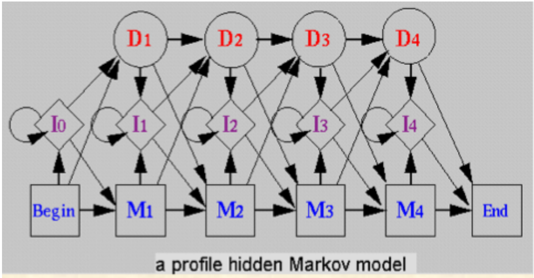


a profile hidden Markov model

We can use a different notation for the HMM.

- $Begin \rightarrow eM_1, \quad e \in \{a, c, g, t\}$
  $Begin \rightarrow eI_0$
  $Begin \rightarrow D_1$
- $M_i \rightarrow eM_{i+1}$

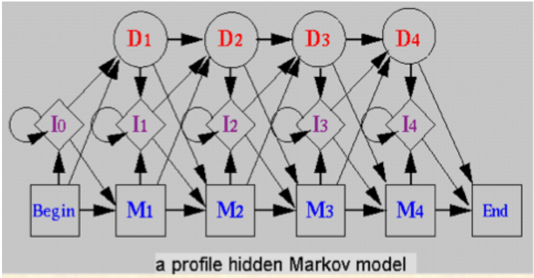
Part IV Probabilistic Models and Learning



We can use a different notation for the HMM.

- $Begin \to eM_1, \quad e \in \{a, c, g, t\}$
  $Begin \to eI_0$
  $Begin \to D_1$
- $M_i \to eM_{i+1}$
  $M_i \to eI_i$

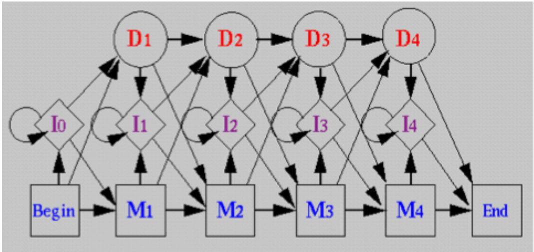
Part IV Probabilistic Models and Learning



We can use a different notation for the HMM.

- $Begin \rightarrow eM_1, \quad e \in \{a, c, g, t\}$
  $Begin \rightarrow eI_0$
  $Begin \rightarrow D_1$
- $M_i \rightarrow eM_{i+1}$
  $M_i \rightarrow eI_i$
  $M_i \rightarrow D_{i+1}$

Part IV Probabilistic Models and Learning



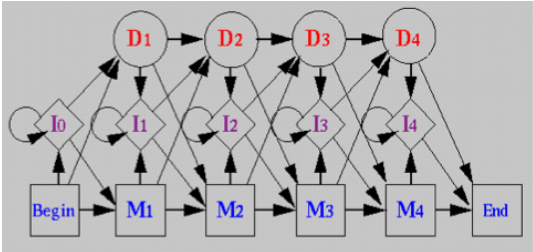
a profile hidden Markov model

We can use a different notation for the HMM.

- $Begin \to eM_1, \quad e \in \{a, c, g, t\}$
  $Begin \to eI_0$
  $Begin \to D_1$

- $M_i \to eM_{i+1}$
  $M_i \to eI_i$
  $M_i \to D_{i+1}$

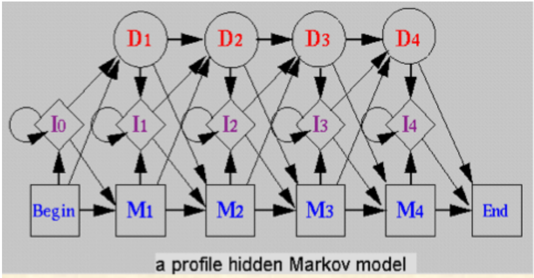
Part IV Probabilistic Models and Learning



a profile hidden Markov model

We can use a different notation for the HMM.

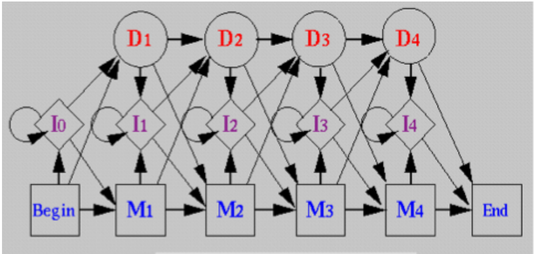
Part IV Probabilistic Models and Learning



We can use a different notation for the HMM.

- $Begin \to aM_1, \; Begin \to cM_1, \; Begin \to gM_1, \; Begin \to tM_1$

Part IV Probabilistic Models and Learning

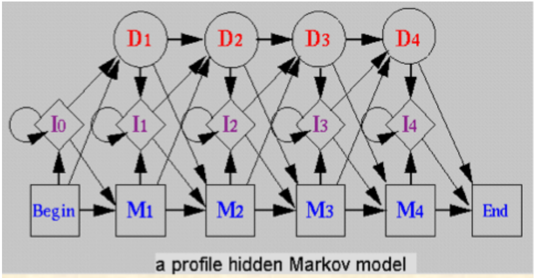


a profile hidden Markov model

We can use a different notation for the HMM.

- $Begin \to aM_1, \; Begin \to cM_1, \; Begin \to gM_1, \; Begin \to tM_1$
  $Begin \to aI_0, \quad Begin \to cI_0, \quad Begin \to gI_0, \quad Begin \to tI_0$

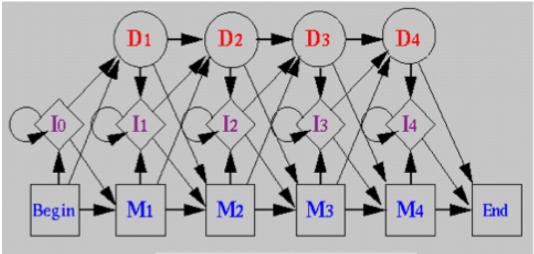
Part IV Probabilistic Models and Learning



We can use a different notation for the HMM.

- $Begin \to aM_1, \ Begin \to cM_1, \ Begin \to gM_1, \ Begin \to tM_1$
  $Begin \to aI_0, \quad Begin \to cI_0, \quad Begin \to gI_0, \quad Begin \to tI_0$
  $Begin \to D_1$

Part IV Probabilistic Models and Learning

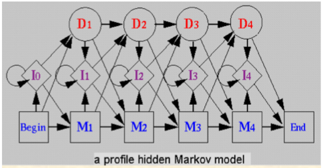


a profile hidden Markov model

We can use a different notation for the HMM.

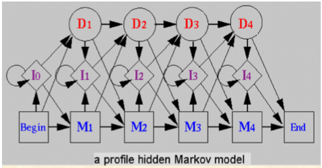
- $Begin \to aM_1,\ Begin \to cM_1,\ Begin \to gM_1,\ Begin \to tM_1$
  $Begin \to aI_0,\ \ Begin \to cI_0,\ \ Begin \to gI_0,\ \ Begin \to tI_0$
  $Begin \to D_1$
- $M_i \to eM_{i+1}$
  $M_i \to eI_i$
  $M_i \to D_{i+1}$

Part IV Probabilistic Models and Learning



sequence $acggt$

Part IV Probabilistic Models and Learning

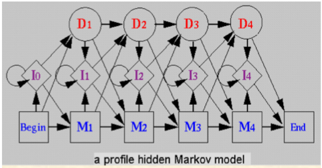


a profile hidden Markov model

sequence $acggt$

- $Begin \to aM_1, \ldots$

Part IV Probabilistic Models and Learning



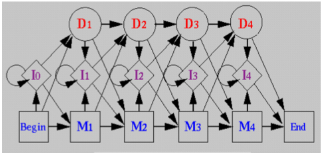
a profile hidden Markov model

sequence $acggt$

- $Begin \to aM_1, \ \ldots$

- $$\begin{array}{llll} M_i \to aM_{i+1}, & M_i \to aI_i, & I_i \to aM_{i+1} & \\ M_i \to cM_{i+1}, & M_i \to cI_i, & I_i \to cM_{i+1} & \\ M_i \to tM_{i+1}, & M_i \to gI_i, & I_i \to gM_{i+1} & \\ M_i \to tM_{i+1}, & M_i \to tI_i, & I_i \to tM_{i+1}, & \text{for } i = 1, 2, 3. \\ M_4 \to End & & & \end{array}$$

Part IV Probabilistic Models and Learning



a profile hidden Markov model

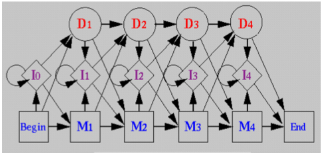
sequence $acggt$

- $Begin \rightarrow aM_1, \ \ldots$

- $M_i \rightarrow aM_{i+1}, \quad M_i \rightarrow aI_i, \quad I_i \rightarrow aM_{i+1}$
  $M_i \rightarrow cM_{i+1}, \quad M_i \rightarrow cI_i, \quad I_i \rightarrow cM_{i+1}$
  $M_i \rightarrow tM_{i+1}, \quad M_i \rightarrow gI_i, \quad I_i \rightarrow gM_{i+1}$
  $M_i \rightarrow tM_{i+1}, \quad M_i \rightarrow tI_i, \quad I_i \rightarrow tM_{i+1},$ for $i = 1, 2, 3.$
  $M_4 \rightarrow End$

we use rules to produce $acggt$:

Part IV Probabilistic Models and Learning



a profile hidden Markov model

sequence $acggt$

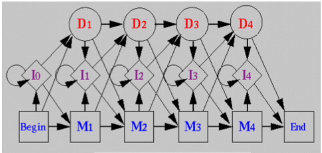
- $Begin \to aM_1, \ \ldots$

- $$\begin{array}{llll} M_i \to aM_{i+1}, & M_i \to aI_i, & I_i \to aM_{i+1} & \\ M_i \to cM_{i+1}, & M_i \to cI_i, & I_i \to cM_{i+1} & \\ M_i \to tM_{i+1}, & M_i \to gI_i, & I_i \to gM_{i+1} & \\ M_i \to tM_{i+1}, & M_i \to tI_i, & I_i \to tM_{i+1}, & \text{for } i = 1, 2, 3. \\ M_4 \to End & & & \end{array}$$

we use rules to produce $acggt$:

$$Begin$$

Part IV Probabilistic Models and Learning



a profile hidden Markov model

sequence $acggt$

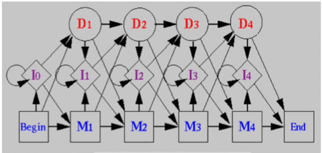
- $Begin \rightarrow aM_1, \ldots$

- $$\begin{array}{llll} M_i \rightarrow aM_{i+1}, & M_i \rightarrow aI_i, & I_i \rightarrow aM_{i+1} & \\ M_i \rightarrow cM_{i+1}, & M_i \rightarrow cI_i, & I_i \rightarrow cM_{i+1} & \\ M_i \rightarrow tM_{i+1}, & M_i \rightarrow gI_i, & I_i \rightarrow gM_{i+1} & \\ M_i \rightarrow tM_{i+1}, & M_i \rightarrow tI_i, & I_i \rightarrow tM_{i+1}, & \text{for } i = 1, 2, 3. \\ M_4 \rightarrow End & & & \end{array}$$

we use rules to produce $acggt$:

$$Begin \Rightarrow aM_1$$

Part IV Probabilistic Models and Learning



a profile hidden Markov model

sequence $acggt$

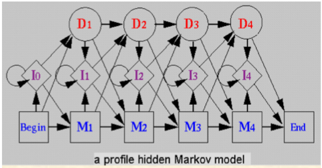
- $Begin \to aM_1, \ldots$

- $M_i \to aM_{i+1}, \quad M_i \to aI_i, \quad I_i \to aM_{i+1}$
  $M_i \to cM_{i+1}, \quad M_i \to cI_i, \quad I_i \to cM_{i+1}$
  $M_i \to tM_{i+1}, \quad M_i \to gI_i, \quad I_i \to gM_{i+1}$
  $M_i \to tM_{i+1}, \quad M_i \to tI_i, \quad I_i \to tM_{i+1}$, for $i = 1, 2, 3.$
  $M_4 \to End$

we use rules to produce $acggt$:

$$Begin \Rightarrow aM_1 \Rightarrow acM_2$$

Part IV Probabilistic Models and Learning



a profile hidden Markov model

sequence $acggt$

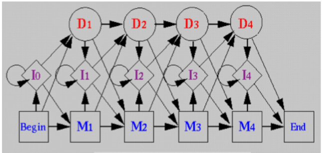
- $Begin \to aM_1, \ldots$

- $$\begin{array}{llll} M_i \to aM_{i+1}, & M_i \to aI_i, & I_i \to aM_{i+1} & \\ M_i \to cM_{i+1}, & M_i \to cI_i, & I_i \to cM_{i+1} & \\ M_i \to tM_{i+1}, & M_i \to gI_i, & I_i \to gM_{i+1} & \\ M_i \to tM_{i+1}, & M_i \to tI_i, & I_i \to tM_{i+1}, & \text{for } i = 1, 2, 3. \\ M_4 \to End & & & \end{array}$$

we use rules to produce $acggt$:

$$Begin \Rightarrow aM_1 \Rightarrow acM_2 \Rightarrow acgI_2$$

Part IV Probabilistic Models and Learning



a profile hidden Markov model

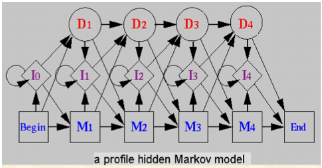
sequence $acggt$

- $Begin \to aM_1, \ldots$
- $$\begin{array}{llll} M_i \to aM_{i+1}, & M_i \to aI_i, & I_i \to aM_{i+1} & \\ M_i \to cM_{i+1}, & M_i \to cI_i, & I_i \to cM_{i+1} & \\ M_i \to tM_{i+1}, & M_i \to gI_i, & I_i \to gM_{i+1} & \\ M_i \to tM_{i+1}, & M_i \to tI_i, & I_i \to tM_{i+1}, & \text{for } i = 1, 2, 3. \\ M_4 \to End & & & \end{array}$$

we use rules to produce $acggt$:

$$Begin \Rightarrow aM_1 \Rightarrow acM_2 \Rightarrow acgI_2 \Rightarrow acggM_3$$

Part IV Probabilistic Models and Learning



a profile hidden Markov model

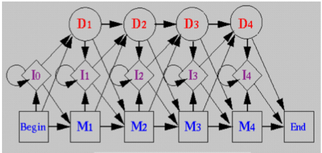
sequence $acggt$

- $Begin \to aM_1, \ldots$
- $$\begin{array}{llll} M_i \to aM_{i+1}, & M_i \to aI_i, & I_i \to aM_{i+1} & \\ M_i \to cM_{i+1}, & M_i \to cI_i, & I_i \to cM_{i+1} & \\ M_i \to tM_{i+1}, & M_i \to gI_i, & I_i \to gM_{i+1} & \\ M_i \to tM_{i+1}, & M_i \to tI_i, & I_i \to tM_{i+1}, & \text{for } i = 1, 2, 3. \\ M_4 \to End & & & \end{array}$$

we use rules to produce $acggt$:

$$Begin \Rightarrow aM_1 \Rightarrow acM_2 \Rightarrow acgI_2 \Rightarrow acggM_3 \Rightarrow acggtM_4$$

Part IV Probabilistic Models and Learning



a profile hidden Markov model

sequence $acggt$

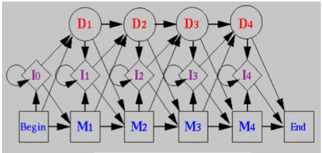
- $Begin \to aM_1, \ldots$

- $M_i \to aM_{i+1}, \quad M_i \to aI_i, \quad I_i \to aM_{i+1}$
  $M_i \to cM_{i+1}, \quad M_i \to cI_i, \quad I_i \to cM_{i+1}$
  $M_i \to tM_{i+1}, \quad M_i \to gI_i, \quad I_i \to gM_{i+1}$
  $M_i \to tM_{i+1}, \quad M_i \to tI_i, \quad I_i \to tM_{i+1},$ for $i = 1, 2, 3.$
  $M_4 \to End$

we use rules to produce $acggt$:

$$Begin \Rightarrow aM_1 \Rightarrow acM_2 \Rightarrow acgI_2 \Rightarrow acggM_3 \Rightarrow acggtM_4 \Rightarrow acggtEnd$$

Part IV Probabilistic Models and Learning



a profile hidden Markov model

sequence $acggt$

- $Begin \rightarrow aM_1, \ldots$

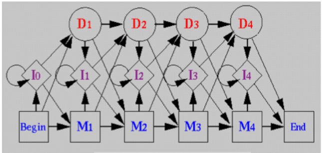
- $$\begin{array}{llll} M_i \rightarrow aM_{i+1}, & M_i \rightarrow aI_i, & I_i \rightarrow aM_{i+1} & \\ M_i \rightarrow cM_{i+1}, & M_i \rightarrow cI_i, & I_i \rightarrow cM_{i+1} & \\ M_i \rightarrow tM_{i+1}, & M_i \rightarrow gI_i, & I_i \rightarrow gM_{i+1} & \\ M_i \rightarrow tM_{i+1}, & M_i \rightarrow tI_i, & I_i \rightarrow tM_{i+1}, & \text{for } i = 1, 2, 3. \\ M_4 \rightarrow End & & & \end{array}$$

we use rules to produce $acggt$:

$$Begin \Rightarrow aM_1 \Rightarrow acM_2 \Rightarrow acgI_2 \Rightarrow acggM_3 \Rightarrow acggtM_4 \Rightarrow acggtEnd$$

Called a **derivation**

Part IV Probabilistic Models and Learning



a profile hidden Markov model

sequence $acggt$

- $Begin \to aM_1, \ldots$

- $$\begin{array}{llll} M_i \to aM_{i+1}, & M_i \to aI_i, & I_i \to aM_{i+1} & \\ M_i \to cM_{i+1}, & M_i \to cI_i, & I_i \to cM_{i+1} & \\ M_i \to tM_{i+1}, & M_i \to gI_i, & I_i \to gM_{i+1} & \\ M_i \to tM_{i+1}, & M_i \to tI_i, & I_i \to tM_{i+1}, & \text{for } i = 1, 2, 3. \\ M_4 \to End & & & \end{array}$$

we use rules to produce $acggt$:

$$Begin \Rightarrow aM_1 \Rightarrow acM_2 \Rightarrow acgI_2 \Rightarrow acggM_3 \Rightarrow acggtM_4 \Rightarrow acggtEnd$$

Called a **derivation** $\Longleftrightarrow$ a **path** in HMM.

Part IV Probabilistic Models and Learning

Notes on the rules:

Part IV Probabilistic Models and Learning

Notes on the rules:

- Rules are *grammar* rules

Notes on the rules:

- Rules are *grammar* rules (also called rewriting rules)

Part IV Probabilistic Models and Learning

Notes on the rules:

- Rules are *grammar* rules (also called rewriting rules)
- Rules like $A \rightarrow aB$ and $A \rightarrow C$ are **regular grammar rules**

Part IV Probabilistic Models and Learning

Notes on the rules:

- Rules are *grammar* rules (also called rewriting rules)
- Rules like $A \rightarrow aB$ and $A \rightarrow C$ are **regular grammar rules** where A, B, C are non-terminals and a is a terminal;

Part IV Probabilistic Models and Learning

Notes on the rules:

- Rules are *grammar* rules (also called rewriting rules)
- Rules like $A \to aB$ and $A \to C$ are **regular grammar rules** where A, B, C are non-terminals and a is a terminal;
- rules can be associated with probability distributions;

Part IV Probabilistic Models and Learning

Notes on the rules:

- Rules are *grammar* rules (also called rewriting rules)
- Rules like $A \rightarrow aB$ and $A \rightarrow C$ are **regular grammar rules** where A, B, C are non-terminals and a is a terminal;
- rules can be associated with probability distributions;
- a derivation is associated with a probability

Part IV Probabilistic Models and Learning

Notes on the rules:

- Rules are *grammar* rules (also called rewriting rules)
- Rules like $A \to aB$ and $A \to C$ are **regular grammar rules** where A, B, C are non-terminals and a is a terminal;
- rules can be associated with probability distributions;
- a derivation is associated with a probability by compounding the probabilities of used rules;

Part IV Probabilistic Models and Learning

Notes on the rules:

- Rules are *grammar* rules (also called rewriting rules)
- Rules like $A \to aB$ and $A \to C$ are **regular grammar rules** where A, B, C are non-terminals and a is a terminal;
- rules can be associated with probability distributions;
- a derivation is associated with a probability by compounding the probabilities of used rules;
- a sequence may have more than one derivation;

Part IV Probabilistic Models and Learning

Notes on the rules:

- Rules are *grammar* rules (also called rewriting rules)
- Rules like $A \rightarrow aB$ and $A \rightarrow C$ are **regular grammar rules** where A, B, C are non-terminals and a is a terminal;
- rules can be associated with probability distributions;
- a derivation is associated with a probability by compounding the probabilities of used rules;
- a sequence may have more than one derivation;
- one of the derivations of the sequence is of the max probability;

Notes on the rules:

- Rules are *grammar* rules (also called rewriting rules)
- Rules like $A \rightarrow aB$ and $A \rightarrow C$ are **regular grammar rules** where A, B, C are non-terminals and a is a terminal;
- rules can be associated with probability distributions;
- a derivation is associated with a probability by compounding the probabilities of used rules;
- a sequence may have more than one derivation;
- one of the derivations of the sequence is of the max probability;
- letters on the sequence are derived one at a time, independently;

Notes on the rules:

- Rules are *grammar* rules (also called rewriting rules)
- Rules like $A \rightarrow aB$ and $A \rightarrow C$ are **regular grammar rules** where A, B, C are non-terminals and a is a terminal;
- rules can be associated with probability distributions;
- a derivation is associated with a probability by compounding the probabilities of used rules;
- a sequence may have more than one derivation;
- one of the derivations of the sequence is of the max probability;
- letters on the sequence are derived one at a time, independently;

- Can rules be designed to model complex relationships among letters?

Part IV Probabilistic Models and Learning

Consider rules for RNA sequences

Part IV Probabilistic Models and Learning

Consider rules for RNA sequences

- $H_i \rightarrow aH_{i+1}u$

Part IV Probabilistic Models and Learning

Consider rules for RNA sequences

- $H_i \rightarrow aH_{i+1}u$
  $H_i \rightarrow uH_{i+1}a$
  $H_i \rightarrow cH_{i+1}g$
  $H_i \rightarrow gH_{i+1}c$

Part IV Probabilistic Models and Learning

Consider rules for RNA sequences

- $H_i \to aH_{i+1}u$
  $H_i \to uH_{i+1}a$
  $H_i \to cH_{i+1}g$
  $H_i \to gH_{i+1}c$
  $H_i \to L_i$

Part IV Probabilistic Models and Learning

Consider rules for RNA sequences

- $H_i \to aH_{i+1}u$
  $H_i \to uH_{i+1}a$
  $H_i \to cH_{i+1}g$
  $H_i \to gH_{i+1}c$
  $H_i \to L_i$

  $L_i \to aL_{i+1}$

Part IV Probabilistic Models and Learning

Consider rules for RNA sequences

- $H_i \to aH_{i+1}u$
  $H_i \to uH_{i+1}a$
  $H_i \to cH_{i+1}g$
  $H_i \to gH_{i+1}c$
  $H_i \to L_i$

  $L_i \to aL_{i+1}$
  $L_i \to cL_{i+1}$
  $L_i \to gL_{i+1}$
  $L_i \to uL_{i+1}$

Part IV Probabilistic Models and Learning

Consider rules for RNA sequences

- $H_i \to aH_{i+1}u$
  $H_i \to uH_{i+1}a$
  $H_i \to cH_{i+1}g$
  $H_i \to gH_{i+1}c$
  $H_i \to L_i$

  $L_i \to aL_{i+1}$
  $L_i \to cL_{i+1}$
  $L_i \to gL_{i+1}$
  $L_i \to uL_{i+1}$
  $L_i \to \epsilon$

Part IV Probabilistic Models and Learning

Consider rules for RNA sequences

- $H_i \to aH_{i+1}u$
  $H_i \to uH_{i+1}a$
  $H_i \to cH_{i+1}g$
  $H_i \to gH_{i+1}c$
  $H_i \to L_i$

  $L_i \to aL_{i+1}$
  $L_i \to cL_{i+1}$
  $L_i \to gL_{i+1}$
  $L_i \to uL_{i+1}$
  $L_i \to \epsilon$ (empty)

Part IV Probabilistic Models and Learning

Consider rules for RNA sequences

- $H_i \to aH_{i+1}u$
  $H_i \to uH_{i+1}a$
  $H_i \to cH_{i+1}g$
  $H_i \to gH_{i+1}c$
  $H_i \to L_i$

  $L_i \to aL_{i+1}$
  $L_i \to cL_{i+1}$
  $L_i \to gL_{i+1}$
  $L_i \to uL_{i+1}$
  $L_i \to \epsilon$ (empty)

- a derivation an RNA sequence that folds into a stem-loop

$$H_0$$

Part IV Probabilistic Models and Learning

Consider rules for RNA sequences

- $H_i \to aH_{i+1}u$
  $H_i \to uH_{i+1}a$
  $H_i \to cH_{i+1}g$
  $H_i \to gH_{i+1}c$
  $H_i \to L_i$

  $L_i \to aL_{i+1}$
  $L_i \to cL_{i+1}$
  $L_i \to gL_{i+1}$
  $L_i \to uL_{i+1}$
  $L_i \to \epsilon$ (empty)

- a derivation an RNA sequence that folds into a stem-loop

$$H_0 \Rightarrow aH_1u$$

Part IV Probabilistic Models and Learning

Consider rules for RNA sequences

- $H_i \rightarrow aH_{i+1}u$
  $H_i \rightarrow uH_{i+1}a$
  $H_i \rightarrow cH_{i+1}g$
  $H_i \rightarrow gH_{i+1}c$
  $H_i \rightarrow L_i$

  $L_i \rightarrow aL_{i+1}$
  $L_i \rightarrow cL_{i+1}$
  $L_i \rightarrow gL_{i+1}$
  $L_i \rightarrow uL_{i+1}$
  $L_i \rightarrow \epsilon$ (empty)

- a derivation an RNA sequence that folds into a stem-loop

$$H_0 \Rightarrow aH_1u \Rightarrow agH_2cu$$

Part IV Probabilistic Models and Learning

Consider rules for RNA sequences

- $H_i \to aH_{i+1}u$
  $H_i \to uH_{i+1}a$
  $H_i \to cH_{i+1}g$
  $H_i \to gH_{i+1}c$
  $H_i \to L_i$

  $L_i \to aL_{i+1}$
  $L_i \to cL_{i+1}$
  $L_i \to gL_{i+1}$
  $L_i \to uL_{i+1}$
  $L_i \to \epsilon$ (empty)

- a derivation an RNA sequence that folds into a stem-loop

$$H_0 \Rightarrow aH_1u \Rightarrow agH_2cu \Rightarrow agaH_3ucu$$

Part IV Probabilistic Models and Learning

Consider rules for RNA sequences

- $H_i \to aH_{i+1}u$
  $H_i \to uH_{i+1}a$
  $H_i \to cH_{i+1}g$
  $H_i \to gH_{i+1}c$
  $H_i \to L_i$

  $L_i \to aL_{i+1}$
  $L_i \to cL_{i+1}$
  $L_i \to gL_{i+1}$
  $L_i \to uL_{i+1}$
  $L_i \to \epsilon$ (empty)

- a derivation an RNA sequence that folds into a stem-loop

$$H_0 \Rightarrow aH_1u \Rightarrow agH_2cu \Rightarrow agaH_3ucu \Rightarrow agaL_3ucu$$

Part IV Probabilistic Models and Learning

Consider rules for RNA sequences

- $H_i \to aH_{i+1}u$
  $H_i \to uH_{i+1}a$
  $H_i \to cH_{i+1}g$
  $H_i \to gH_{i+1}c$
  $H_i \to L_i$

  $L_i \to aL_{i+1}$
  $L_i \to cL_{i+1}$
  $L_i \to gL_{i+1}$
  $L_i \to uL_{i+1}$
  $L_i \to \epsilon$ (empty)

- a derivation an RNA sequence that folds into a stem-loop

$$H_0 \Rightarrow aH_1u \Rightarrow agH_2cu \Rightarrow agaH_3ucu \Rightarrow agaL_3ucu$$

$$\Rightarrow agaaL_4ucu$$

Part IV Probabilistic Models and Learning

Consider rules for RNA sequences

- $H_i \to aH_{i+1}u$
  $H_i \to uH_{i+1}a$
  $H_i \to cH_{i+1}g$
  $H_i \to gH_{i+1}c$
  $H_i \to L_i$

  $L_i \to aL_{i+1}$
  $L_i \to cL_{i+1}$
  $L_i \to gL_{i+1}$
  $L_i \to uL_{i+1}$
  $L_i \to \epsilon$ (empty)

- a derivation an RNA sequence that folds into a stem-loop

$$H_0 \Rightarrow aH_1u \Rightarrow agH_2cu \Rightarrow agaH_3ucu \Rightarrow agaL_3ucu$$

$$\Rightarrow agaaL_4ucu \Rightarrow agaaaL_5ucu$$

Part IV Probabilistic Models and Learning

Consider rules for RNA sequences

- $H_i \to aH_{i+1}u$
  $H_i \to uH_{i+1}a$
  $H_i \to cH_{i+1}g$
  $H_i \to gH_{i+1}c$
  $H_i \to L_i$

  $L_i \to aL_{i+1}$
  $L_i \to cL_{i+1}$
  $L_i \to gL_{i+1}$
  $L_i \to uL_{i+1}$
  $L_i \to \epsilon$ (empty)

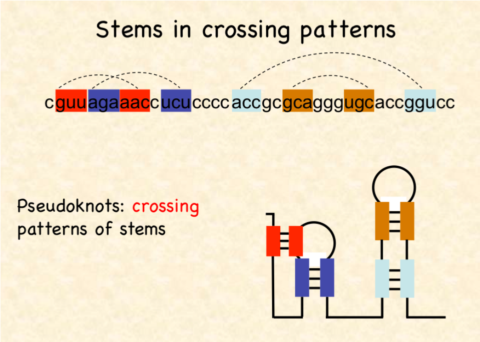
- a derivation an RNA sequence that folds into a stem-loop

$$H_0 \Rightarrow aH_1u \Rightarrow agH_2cu \Rightarrow agaH_3ucu \Rightarrow agaL_3ucu$$

$$\Rightarrow agaaL_4ucu \Rightarrow agaaaL_5ucu \Rightarrow agaaaucu$$

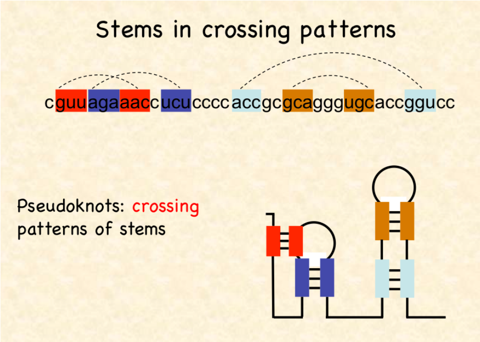
Part IV Probabilistic Models and Learning

RNA secondary structure examples:



Part IV Probabilistic Models and Learning

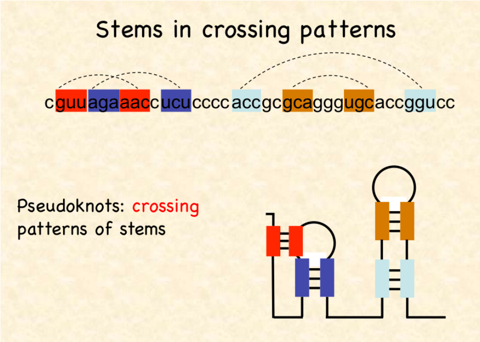
RNA secondary structure examples:



- nesting, parallel patterns are context-free, while

Part IV Probabilistic Models and Learning

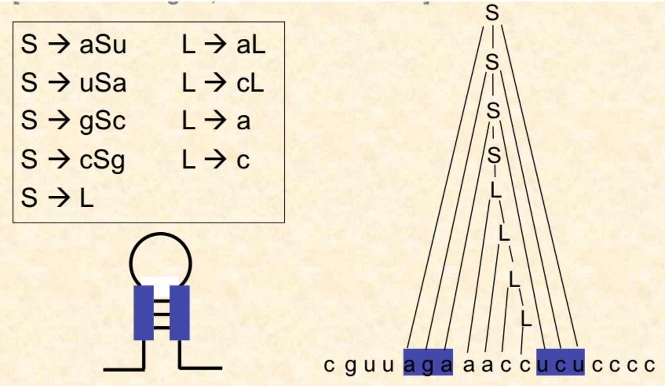
RNA secondary structure examples:



- nesting, parallel patterns are context-free, while
- crossing patterns are not!

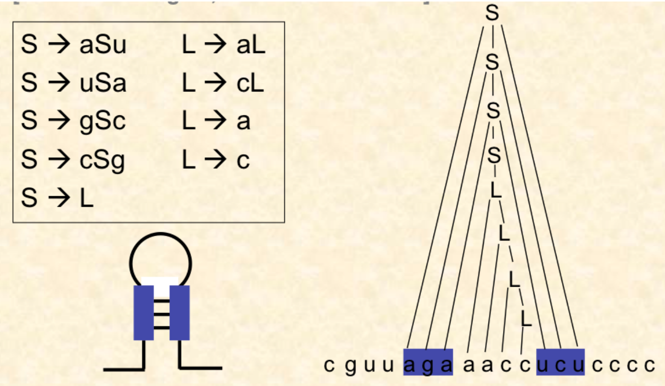
Part IV Probabilistic Models and Learning

Illustration of context-free grammar derivation:



Part IV Probabilistic Models and Learning

Illustration of context-free grammar derivation:

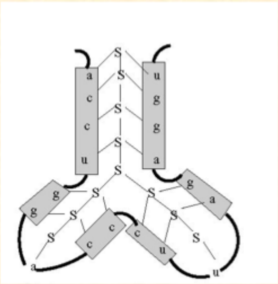


- Context-free grammar derivation is a tree (because of simultaneous emissions)

Part IV Probabilistic Models and Learning

Illustration of CFG derivation again:

1. A CFG	2. A derivation of "accuggacccuuagaggu"
S → aSu	S → aSu
S → cSg	→ acSgu
S → gSc	→ accSggu
S → uSa	→ accuSaggu
S → a	→ accuSSaggu
S → c	→ accugScSaggu
S → g	→ accuggSccSaggu
S → u	→ accuggaccSaggu
S → SS	→ accuggacccSgaggu
	→ accuggaccuSagaggu
	→ accuggacccuuagaggu



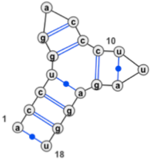
1. A CFG   2. A derivation of "accuggacccuuagaggu"   3. Corresponding structure

Part IV Probabilistic Models and Learning

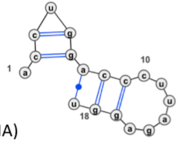
S → aSu
S → cSg
S → gSc
S → uSa
S → a
S → c
S → g
S → u
S → SS

S → aSu
→ acSgu
→ accSggu
→ accuSaggu
→ accuSSaggu
→ accugScSaggu
→ accuggSccSaggu
→ accuggaccSaggu
→ accuggacccSgaggu
→ accuggacccuSagaggu
→ accuggacccuuagaggu

S → SS
→ SSS
→ aSS
→ acSgS
→ accSggS
→ accuggS
→ accuggaSu
→ accuggacSgu
→ accuggaccSggu
→ ......
→ accuggacccuuagaggu



(Drawn with VARNA)



(A CFG applied on the same sequence with two alternative syntactic structures)

Part IV Probabilistic Models and Learning

Stochastic context-free grammar (SCFG):

- Probability distributions are associated with grammar rules

1. $S \to aSb$	$\{0.4\}$	4. $S \to a$	$\{0.1\}$
2. $S \to aS$	$\{0.1\}$	5. $S \to b$	$\{0.1\}$
3. $S \to bS$	$\{0.1\}$	6. $S \to SS$	$\{0.2\}$

Part IV Probabilistic Models and Learning

Stochastic context-free grammar (SCFG):

- Probability distributions are associated with grammar rules

1. $S \rightarrow aSb$	{0.4}	4. $S \rightarrow a$	{0.1}
2. $S \rightarrow aS$	{0.1}	5. $S \rightarrow b$	{0.1}
3. $S \rightarrow bS$	{0.1}	6. $S \rightarrow SS$	{0.2}

for every variable X, $\sum_{X \rightarrow \alpha} Prob(X \rightarrow \alpha) = 1$

Part IV Probabilistic Models and Learning

Stochastic context-free grammar (SCFG):

- Probability distributions are associated with grammar rules

1. $S \to aSb$	$\{0.4\}$	4. $S \to a$	$\{0.1\}$
2. $S \to aS$	$\{0.1\}$	5. $S \to b$	$\{0.1\}$
3. $S \to bS$	$\{0.1\}$	6. $S \to SS$	$\{0.2\}$

for every variable X, $\sum_{X \to \alpha} Prob(X \to \alpha) = 1$

- Every syntax structure of a sequence is associated with a probability.

Stochastic context-free grammar (SCFG):

- Probability distributions are associated with grammar rules

1. $S \to aSb$	{0.4}	4. $S \to a$	{0.1}
2. $S \to aS$	{0.1}	5. $S \to b$	{0.1}
3. $S \to bS$	{0.1}	6. $S \to SS$	{0.2}

for every variable X, $\sum_{X \to \alpha} Prob(X \to \alpha) = 1$

- Every syntax structure of a sequence is associated with a probability.

π_A: $\underline{S} \Rightarrow_1 a\underline{S}b \Rightarrow_1 aa\underline{S}bb \Rightarrow_3 aab\underline{S}bb \Rightarrow_4 aababb = x$
π_B: $\underline{S} \Rightarrow_6 \underline{S}S \Rightarrow_1 a\underline{S}bS \Rightarrow_4 aab\underline{S} \Rightarrow_1 aaba\underline{S}b \Rightarrow_5 aababb = x$

Part IV Probabilistic Models and Learning

Stochastic context-free grammar (SCFG):

- Probability distributions are associated with grammar rules

1. $S \to aSb$	$\{0.4\}$	4. $S \to a$	$\{0.1\}$
2. $S \to aS$	$\{0.1\}$	5. $S \to b$	$\{0.1\}$
3. $S \to bS$	$\{0.1\}$	6. $S \to SS$	$\{0.2\}$

for every variable X, $\sum_{X \to \alpha} Prob(X \to \alpha) = 1$

- Every syntax structure of a sequence is associated with a probability.

π_A: $\underline{S} \Rightarrow_1 a\underline{S}b \Rightarrow_1 aa\underline{S}bb \Rightarrow_3 aab\underline{S}bb \Rightarrow_4 aababb = x$

π_B: $\underline{S} \Rightarrow_6 \underline{SS} \Rightarrow_1 a\underline{S}bS \Rightarrow_4 aab\underline{S} \Rightarrow_1 aaba\underline{S}b \Rightarrow_5 aababb = x$

$$Prob(\pi_A, x) = 0.4 \times 0.4 \times 0.1 \times 0.1 = 0.016$$
$$Prob(\pi_B, x) = 0.2 \times 0.4 \times 0.1 \times 0.4 \times 0.1 = 0.0032$$

Part IV Probabilistic Models and Learning

RNA secondary structure modeling with SCFG:

RNA secondary structure modeling with SCFG:

- Effective
  - specific enough for profiling
  - general enough for structure prediction

Part IV Probabilistic Models and Learning

RNA secondary structure modeling with SCFG:

- Effective
  - specific enough for profiling
  - general enough for structure prediction
- Efficient: $O(n^3)$-time computations
  - decoding (structure prediction)
  - structure analysis (structural alignment)
  - probability parameter estimation

Part IV Probabilistic Models and Learning

RNA secondary structure modeling with SCFG:

- Effective
  - specific enough for profiling
  - general enough for structure prediction
- Efficient: $O(n^3)$-time computations
  - decoding (structure prediction)
  - structure analysis (structural alignment)
  - probability parameter estimation
- performance
  - comparable to energy-based methods
  - unique and successful in structural profile-based search

Part IV Probabilistic Models and Learning

RNA secondary structure modeling with SCFG:

- Effective
  - specific enough for profiling
  - general enough for structure prediction
- Efficient: $O(n^3)$-time computations
  - decoding (structure prediction)
  - structure analysis (structural alignment)
  - probability parameter estimation
- performance
  - comparable to energy-based methods
  - unique and successful in structural profile-based search

[Sakakibara et al, 1994, Eddy and Durbin 1994,
 . . . ,
Rivas et al, 2012]

Part IV Probabilistic Models and Learning

Stochastic grammars for (RNA) tertiary structure modeling?

Part IV Probabilistic Models and Learning

Stochastic grammars for (RNA) tertiary structure modeling?

- much smaller set of resolved 3D structures
  (in contrast to proteins or reported RNA secondary structures)

Stochastic grammars for (RNA) tertiary structure modeling?

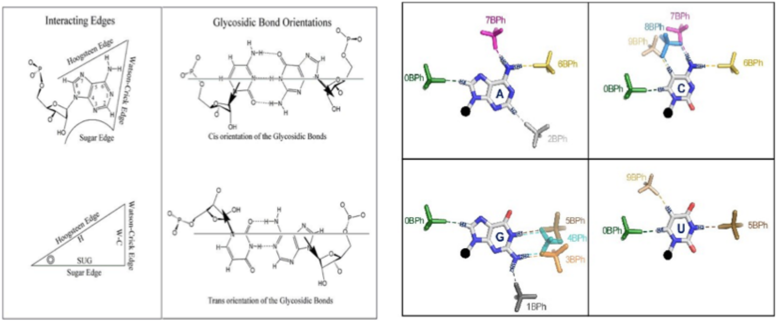
- much smaller set of resolved 3D structures
  (in contrast to proteins or reported RNA secondary structures)
- tertiary interactions were not understood until recently

Part IV Probabilistic Models and Learning

Stochastic grammars for (RNA) tertiary structure modeling?

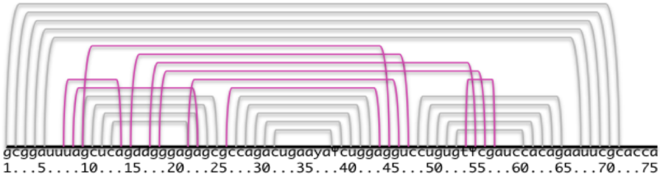
- much smaller set of resolved 3D structures
  (in contrast to proteins or reported RNA secondary structures)

- tertiary interactions were not understood until recently



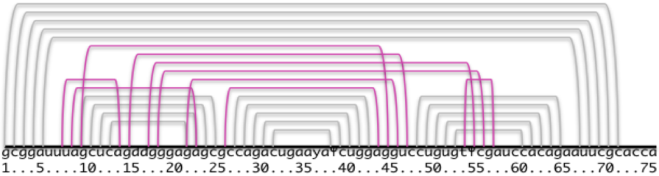
(Leontis et al, 2003; Zirbel et al, 2009. 12 base-base, 10 base-phosphate, and 10 base-ribose families)

All nucleotide interactions of a tRNA (excluding stacking)



Part IV Probabilistic Models and Learning

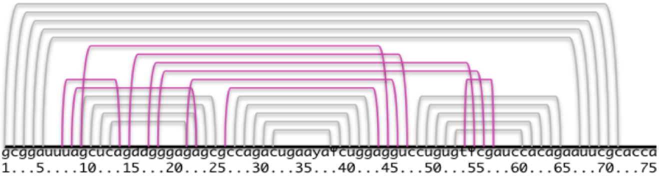
All nucleotide interactions of a tRNA (excluding stacking)



- gray relation is context-free;

Part IV Probabilistic Models and Learning

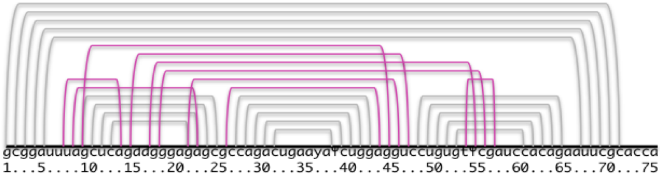
All nucleotide interactions of a tRNA (excluding stacking)



- gray relation is context-free;
- purple relation is context-sensitive.

Part IV Probabilistic Models and Learning

All nucleotide interactions of a tRNA (excluding stacking)



- gray relation is context-free;
- purple relation is context-sensitive.
- We need a higher-order model for such complex relations!

Part IV Probabilistic Models and Learning

Part IV Probabilistic Models and Learning

3. Markov networks and learning

3. Markov networks and learning

- Compute joint probability distribution $P(X)$ from observed random variables $X = \langle X_1, \ldots, X_n \rangle$

  Example 1: molecule residues forming structure

3. Markov networks and learning

- Compute joint probability distribution $P(X)$ from observed random variables $X = \langle X_1, \ldots, X_n \rangle$

  Example 1: molecule residues forming structure

  Example 2: gene networks from expression data

Part IV Probabilistic Models and Learning

Part IV Probabilistic Models and Learning

- $P(X)$ is a n^{th} order distribution, difficult to compute

Part IV Probabilistic Models and Learning

- $P(X)$ is a n^{th} order distribution, difficult to compute
- Approximation with a second order distribution $P_G(X)$ (i.e., binary relation, Markov network)

Part IV Probabilistic Models and Learning

- $P(X)$ is a n^{th} order distribution, difficult to compute
- Approximation with a second order distribution $P_G(X)$ (i.e., binary relation, Markov network)

  e.g., molecule residues are random variables

Part IV Probabilistic Models and Learning

- $P(X)$ is a n^{th} order distribution, difficult to compute
- Approximation with a second order distribution $P_G(X)$ (i.e., binary relation, Markov network)

  e.g., molecule residues are random variables

    molecular structure is defined over their joint distribution, involving multi-body interactions.

    Markov network model approximates multi-body interactions with pairwise interactions.

Part IV Probabilistic Models and Learning

Part IV Probabilistic Models and Learning

Questions to answer:

Part IV Probabilistic Models and Learning

Questions to answer:

- What does $P_G(X)$ look like

Part IV Probabilistic Models and Learning

Questions to answer:

- What does $P_G(X)$ look like even a Markov graph G is given?

Part IV Probabilistic Models and Learning

Questions to answer:

- What does $P_G(X)$ look like even a Markov graph G is given?
- How to measure the difference between $P_G(X)$ and $P(X)$?

Part IV Probabilistic Models and Learning

Questions to answer:

- What does $P_G(X)$ look like even a Markov graph G is given?
- How to measure the difference between $P_G(X)$ and $P(X)$?
- Can we compute G and $P_G(X)$ efficiently?

Part IV Probabilistic Models and Learning

The framework of Chow and Liu 1968:

Part IV Probabilistic Models and Learning

The framework of Chow and Liu 1968:

The framework of Chow and Liu 1968:

- to measure difference between two distributions $P(X)$ and $P_G(X)$) with D_{KL}, Kullback-Leibler divergence;

Part IV Probabilistic Models and Learning

The framework of Chow and Liu 1968:

- to measure difference between two distributions $P(X)$ and $P_G(X))$ with D_{KL}, Kullback-Leibler divergence;
- when G is assumed to be of tree topology,

Part IV Probabilistic Models and Learning

The framework of Chow and Liu 1968:

- to measure difference between two distributions $P(X)$ and $P_G(X))$ with D_{KL}, Kullback-Leibler divergence;
- when G is assumed to be of tree topology, minimizing D_{KL} results in maximum spanning tree problem

The framework of Chow and Liu 1968:

- to measure difference between two distributions $P(X)$ and $P_G(X))$ with D_{KL}, Kullback-Leibler divergence;
- when G is assumed to be of tree topology, minimizing D_{KL} results in maximum spanning tree problem
- If non-tree topology is desired

The framework of Chow and Liu 1968:

- to measure difference between two distributions $P(X)$ and $P_G(X))$ with D_{KL}, Kullback-Leibler divergence;
- when G is assumed to be of tree topology, minimizing D_{KL} results in maximum spanning tree problem
- If non-tree topology is desired
  1. the problem becomes computationally intractable;

The framework of Chow and Liu 1968:

- to measure difference between two distributions $P(X)$ and $P_G(X))$ with D_{KL}, Kullback-Leibler divergence;
- when G is assumed to be of tree topology, minimizing D_{KL} results in maximum spanning tree problem
- If non-tree topology is desired
  1. the problem becomes computationally intractable;
  2. relying on heuristics.

Part IV Probabilistic Models and Learning

Assume we have a Markov tree T for variable $X = \{X_1, \ldots, X_n\}$ with with a root X_1

Part IV Probabilistic Models and Learning

Assume we have a Markov tree T for variable $X = \{X_1, \ldots, X_n\}$ with with a root X_1

Part IV Probabilistic Models and Learning

Assume we have a Markov tree T for variable $X = \{X_1, \ldots, X_n\}$ with with a root X_1

- tree topology is completely determined by π, the parent information e.g., $\pi(6) = 3$

Assume we have a Markov tree T for variable $X = \{X_1, \ldots, X_n\}$ with with a root X_1

- tree topology is completely determined by π, the parent information e.g., $\pi(6) = 3$
- $P_T(X) = P(X_1) \prod_{i=2}^{n} P(X_i | X_{\pi(i)})$

Part IV Probabilistic Models and Learning

Assume we have a Markov tree T for variable $X = \{X_1, \dots, X_n\}$ with with a root X_1

- tree topology is completely determined by π, the parent information e.g., $\pi(6) = 3$
- $P_T(X) = P(X_1) \prod_{i=2}^{n} P(X_i | X_{\pi(i)})$
- Minimizing $D_{KL}(P(X), P_T(X))$ would tell us what T should be.

Part IV Probabilistic Models and Learning

Kullback-Leilber divergence:

$$D_{KL}((P(X), P_T(X)) = \sum_x P(x) \log_2 \frac{P(x)}{P_T(x)}$$

Part IV Probabilistic Models and Learning

Kullback-Leilber divergence:

$$D_{KL}((P(X), P_T(X)) = \sum_x P(x) \log_2 \frac{P(x)}{P_T(x)}$$

where $x = (x_1, \ldots, x_n)$ is the vector of values for variables $X_1, \ldots, X_n$.

Part IV Probabilistic Models and Learning

Kullback-Leilber divergence:

$$D_{KL}((P(X), P_T(X)) = \sum_x P(x) \log_2 \frac{P(x)}{P_T(x)}$$

where $x = (x_1, \ldots, x_n)$ is the vector of values for variables $X_1, \ldots, X_n$.

$$D_{KL}((P(X), P_T(X)) = \sum_x P(x) \log_2 P(x) - \sum_x P(x) \log_2 P_T(x)$$

Part IV Probabilistic Models and Learning

Kullback-Leilber divergence:

$$D_{KL}((P(X), P_T(X)) = \sum_x P(x) \log_2 \frac{P(x)}{P_T(x)}$$

where $x = (x_1, \ldots, x_n)$ is the vector of values for variables $X_1, \ldots, X_n$.

$$D_{KL}((P(X), P_T(X)) = \sum_x P(x) \log_2 P(x) - \sum_x P(x) \log_2 P_T(x)$$

$$= -H(X) - \sum_x P(x) \log_2 P(x_1) \prod_{i=2}^{n} P(x_i | x_{\pi(i)})$$

Part IV Probabilistic Models and Learning

Kullback-Leilber divergence:

$$D_{KL}((P(X), P_T(X)) = \sum_x P(x) \log_2 \frac{P(x)}{P_T(x)}$$

where $x = (x_1, \dots, x_n)$ is the vector of values for variables $X_1, \dots, X_n$.

$$D_{KL}((P(X), P_T(X)) = \sum_x P(x) \log_2 P(x) - \sum_x P(x) \log_2 P_T(x)$$

$$= -H(X) - \sum_x P(x) \log_2 P(x_1) \prod_{i=2}^{n} P(x_i | x_{\pi(i)})$$

The second term is

$$-\sum_{(x_1,\dots,x_n)} P(x_1, \dots, x_n) \log_2 P(x_1) - \sum_{(x_1,\dots,x_n)} P(x_1, \dots, x_n) \log_2 \prod_{i=2}^{n} P(x_i | X_{\pi(i)})$$

Part IV Probabilistic Models and Learning

Kullback-Leilber divergence:

$$D_{KL}((P(X), P_T(X)) = \sum_x P(x) \log_2 \frac{P(x)}{P_T(x)}$$

where $x = (x_1, \ldots, x_n)$ is the vector of values for variables $X_1, \ldots, X_n$.

$$D_{KL}((P(X), P_T(X)) = \sum_x P(x) \log_2 P(x) - \sum_x P(x) \log_2 P_T(x)$$

$$= -H(X) - \sum_x P(x) \log_2 P(x_1) \prod_{i=2}^{n} P(x_i | x_{\pi(i)})$$

The second term is

$$- \sum_{(x_1,\ldots,x_n)} P(x_1, \ldots, x_n) \log_2 P(x_1) - \sum_{(x_1,\ldots,x_n)} P(x_1, \ldots, x_n) \log_2 \prod_{i=2}^{n} P(x_i | X_{\pi(i)})$$

$$= -\sum_{x_1} P(x_1) \log_2 P(x_1) - \sum_{(x_1,\ldots,x_n)} P(x_1, \ldots, x_n) \sum_{i=2}^{n} \log_2 P(x_i | x_{\pi(i)})$$

Part IV Probabilistic Models and Learning

Kullback-Leilber divergence:

$$D_{KL}((P(X), P_T(X)) = \sum_x P(x) \log_2 \frac{P(x)}{P_T(x)}$$

where $x = (x_1, \ldots, x_n)$ is the vector of values for variables $X_1, \ldots, X_n$.

$$D_{KL}((P(X), P_T(X)) = \sum_x P(x) \log_2 P(x) - \sum_x P(x) \log_2 P_T(x)$$

$$= -H(X) - \sum_x P(x) \log_2 P(x_1) \prod_{i=2}^{n} P(x_i | x_{\pi(i)})$$

The second term is

$$-\sum_{(x_1,\ldots,x_n)} P(x_1, \ldots, x_n) \log_2 P(x_1) - \sum_{(x_1,\ldots,x_n)} P(x_1, \ldots, x_n) \log_2 \prod_{i=2}^{n} P(x_i | X_{\pi(i)})$$

$$= -\sum_{x_1} P(x_1) \log_2 P(x_1) - \sum_{(x_1,\ldots,x_n)} P(x_1, \ldots, x_n) \sum_{i=2}^{n} \log_2 P(x_i | x_{\pi(i)})$$

$$= H(X_1) - \sum_{(x_1,\ldots,x_n)} P(x_1, \ldots, x_n) \sum_{i=2}^{n} \log_2 P(x_i | x_{\pi(i)})$$

Part IV Probabilistic Models and Learning

Kullback-Leilber divergence:

$$D_{KL}((P(X), P_T(X)) = \sum_x P(x) \log_2 \frac{P(x)}{P_T(x)}$$

where $x = (x_1, \ldots, x_n)$ is the vector of values for variables $X_1, \ldots, X_n$.

$$D_{KL}((P(X), P_T(X)) = \sum_x P(x) \log_2 P(x) - \sum_x P(x) \log_2 P_T(x)$$

$$= -H(X) - \sum_x P(x) \log_2 P(x_1) \prod_{i=2}^{n} P(x_i | x_{\pi(i)})$$

The second term is

$$- \sum_{(x_1,\ldots,x_n)} P(x_1, \ldots, x_n) \log_2 P(x_1) - \sum_{(x_1,\ldots,x_n)} P(x_1, \ldots, x_n) \log_2 \prod_{i=2}^{n} P(x_i | X_{\pi(i)})$$

$$= -\sum_{x_1} P(x_1) \log_2 P(x_1) - \sum_{(x_1,\ldots,x_n)} P(x_1, \ldots, x_n) \sum_{i=2}^{n} \log_2 P(x_i | x_{\pi(i)})$$

$$= H(X_1) - \sum_{(x_1,\ldots,x_n)} P(x_1, \ldots, x_n) \sum_{i=2}^{n} \log_2 P(x_i | x_{\pi(i)})$$

Part IV Probabilistic Models and Learning

from the last slide:

$$= H(X_1) - \sum_{(x_1,\ldots,x_n)} P(x_1,\ldots,x_n) \sum_{i=2}^{n} \log_2 P(x_i|x_{\pi(i)})$$

Part IV Probabilistic Models and Learning

from the last slide:

$$= H(X_1) - \sum_{(x_1,\ldots,x_n)} P(x_1,\ldots,x_n) \sum_{i=2}^{n} \log_2 P(x_i|x_{\pi(i)})$$

$$= H(X_1) - \sum_{(x_1,\ldots,x_n)} P(x_1,\ldots,x_n) \sum_{i=2}^{n} \log_2 P(x_i) \frac{P(x_i|x_{\pi(i)}) P(x_{\pi(i)})}{P(x_i) P(x_{\pi(i)})}$$

Part IV Probabilistic Models and Learning

from the last slide:

$$= H(X_1) - \sum_{(x_1,\ldots,x_n)} P(x_1,\ldots,x_n) \sum_{i=2}^{n} \log_2 P(x_i|x_{\pi(i)})$$

$$= H(X_1) - \sum_{(x_1,\ldots,x_n)} P(x_1,\ldots,x_n) \sum_{i=2}^{n} \log_2 P(x_i) \frac{P(x_i|x_{\pi(i)})P(x_{\pi(i)})}{P(x_i)P(x_{\pi(i)})}$$

$$= H(X_1) - \sum_{i=2}^{n} \sum_{(x_1,\ldots,x_n)} P(x_1,\ldots,x_n) \log_2 P(x_i) \frac{P(x_i, x_{\pi(i)})}{P(x_i)P(x_{\pi(i)})}$$

continued from the previous page

$$= H(X_1) - \sum_{i=2}^{n} \sum_{(x_1,\ldots,x_n)} P(x_1,\ldots,x_n) \log_2 {\color{red}P(x_i)} \frac{P(x_i, x_{\pi(i)})}{P(x_i)P(x_{\pi(i)})}$$

$$= H(X_1) - \sum_{i=2}^{n} \sum_{(x_1,\ldots,x_n)} P(x_1,\ldots,x_n) \log_2 P(x_i)$$
$$- \sum_{i=2}^{n} \sum_{(x_1,\ldots,x_n)} P(x_1,\ldots,x_n) \log_2 \frac{P(x_i, x_{\pi(i)})}{P(x_i)P(x_{\pi(i)})}$$

Part IV Probabilistic Models and Learning

continued from the previous page

$$= H(X_1) - \sum_{i=2}^{n} \sum_{(x_1,\ldots,x_n)} P(x_1,\ldots,x_n) \log_2 P(x_i) \frac{P(x_i, x_{\pi(i)})}{P(x_i)P(x_{\pi(i)})}$$

$$= H(X_1) - \sum_{i=2}^{n} \sum_{(x_1,\ldots,x_n)} P(x_1,\ldots,x_n) \log_2 P(x_i)$$

$$- \sum_{i=2}^{n} \sum_{(x_1,\ldots,x_n)} P(x_1,\ldots,x_n) \log_2 \frac{P(x_i, x_{\pi(i)})}{P(x_i)P(x_{\pi(i)})}$$

$$= H(X_1) - \sum_{i=2}^{n} \sum_{x_i} P(x_i) \log_2 P(x_i) - \sum_{i=2}^{n} \sum_{x_i, x_{\pi(i)}} P(x_i, x_{\pi(i)}) \log_2 \frac{P(x_i, x_{\pi(i)})}{P(x_i)P(x_{\pi(i)})}$$

Part IV Probabilistic Models and Learning

continued from the previous page

$$= H(X_1) - \sum_{i=2}^{n} \sum_{(x_1,\ldots,x_n)} P(x_1,\ldots,x_n) \log_2 {\color{red} P(x_i)} \frac{P(x_i, x_{\pi(i)})}{P(x_i)P(x_{\pi(i)})}$$

$$= H(X_1) - \sum_{i=2}^{n} \sum_{(x_1,\ldots,x_n)} P(x_1,\ldots,x_n) \log_2 P(x_i)$$

$$- \sum_{i=2}^{n} \sum_{(x_1,\ldots,x_n)} P(x_1,\ldots,x_n) \log_2 \frac{P(x_i, x_{\pi(i)})}{P(x_i)P(x_{\pi(i)})}$$

$$= H(X_1) - \sum_{i=2}^{n} \sum_{x_i} P(x_i) \log_2 P(x_i) - \sum_{i=2}^{n} \sum_{x_i, x_{\pi(i)}} P(x_i, x_{\pi(i)}) \log_2 \frac{P(x_i, x_{\pi(i)})}{P(x_i)P(x_{\pi(i)})}$$

$$= H(X_1) + \sum_{i=2}^{n} H(X_i) - \sum_{i=2}^{n} I(X_i, X_{\pi(i)})$$

continued from the previous page

$$= H(X_1) - \sum_{i=2}^{n} \sum_{(x_1,\ldots,x_n)} P(x_1,\ldots,x_n) \log_2 P(x_i) \frac{P(x_i, x_{\pi(i)})}{P(x_i)P(x_{\pi(i)})}$$

$$= H(X_1) - \sum_{i=2}^{n} \sum_{(x_1,\ldots,x_n)} P(x_1,\ldots,x_n) \log_2 P(x_i)$$

$$- \sum_{i=2}^{n} \sum_{(x_1,\ldots,x_n)} P(x_1,\ldots,x_n) \log_2 \frac{P(x_i, x_{\pi(i)})}{P(x_i)P(x_{\pi(i)})}$$

$$= H(X_1) - \sum_{i=2}^{n} \sum_{x_i} P(x_i) \log_2 P(x_i) - \sum_{i=2}^{n} \sum_{x_i, x_{\pi(i)}} P(x_i, x_{\pi(i)}) \log_2 \frac{P(x_i, x_{\pi(i)})}{P(x_i)P(x_{\pi(i)})}$$

$$= H(X_1) + \sum_{i=2}^{n} H(X_i) - \sum_{i=2}^{n} I(X_i, X_{\pi(i)})$$

$$= \sum_{i=1}^{n} H(X_i) - \sum_{i=2}^{n} I(X_i, X_{\pi(i)})$$

Part IV Probabilistic Models and Learning

So Kullback-Leilber divergence:

$$D_{KL}((P(X), P_T(X)) = -H(X) + \sum_{i=1}^{n} H(X_i) - \sum_{i=2}^{n} I(X_i, X_{\pi(i)})$$

So Kullback-Leilber divergence:

$$D_{KL}((P(X), P_T(X)) = -H(X) + \sum_{i=1}^{n} H(X_i) - \sum_{i=2}^{n} I(X_i, X_{\pi(i)})$$

- The left-hand-side is minimized if $\sum_{i=2}^{n} I(X_i, X_{\pi(i)})$ is maximized,

Part IV Probabilistic Models and Learning

So Kullback-Leilber divergence:

$$D_{KL}((P(X), P_T(X)) = -H(X) + \sum_{i=1}^{n} H(X_i) - \sum_{i=2}^{n} I(X_i, X_{\pi(i)})$$

- The left-hand-side is minimized if $\sum_{i=2}^{n} I(X_i, X_{\pi(i)})$ is maximized,
- $I(X_i, X_{\pi(i)}) = \sum_{x_i, x_{\pi(i)}} p(x_i) \log \frac{P(x_i, x_{\pi(i)})}{p(x_i)p(x_{\pi(i)})}$
  is the **mutual information** between X_i and $X_{\pi(i)}$.

Such Markovtree T can be found with the following steps:

Such Markovtree T can be found with the following steps:

- Construct graph G_X of n vertices, one for each variable $X_i \in X$;

Part IV Probabilistic Models and Learning

Such Markovtree T can be found with the following steps:

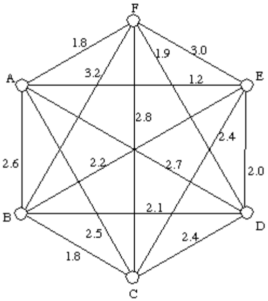
- Construct graph G_X of n vertices, one for each variable $X_i \in X$;
- edge (i, j) has weight $I(X_i, X_j)$, for every pair of i, j;

Such Markovtree T can be found with the following steps:

- Construct graph G_X of n vertices, one for each variable $X_i \in X$;
- edge (i, j) has weight $I(X_i, X_j)$, for every pair of i, j;
- find a maximum spanning tree T of G_X;
  (max spanning tree has the same algorithm as min spanning tree)

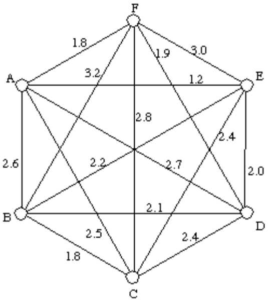
Part IV Probabilistic Models and Learning

Formulated complete graph:

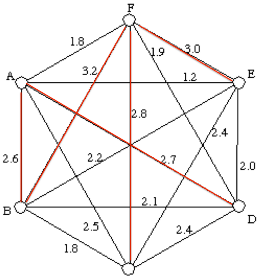


Part IV Probabilistic Models and Learning

Formulated complete graph:



A maximum spanning tree:



We note that:

Part IV Probabilistic Models and Learning

We note that:

- Finding the best Markov tree equals the maximum spanning tree problem;

Part IV Probabilistic Models and Learning

We note that:

- Finding the best Markov tree equals the maximum spanning tree problem;
- Algorithms for MinST suit MaxST, e.g., Prim's, Kruskal's ;

Part IV Probabilistic Models and Learning

We note that:

- Finding the best Markov tree equals the maximum spanning tree problem;
- Algorithms for MinST suit MaxST, e.g., Prim's, Kruskal's ;
- The obtained Markov tree T is not a causation relation,

We note that:

- Finding the best Markov tree equals the maximum spanning tree problem;
- Algorithms for MinST suit MaxST, e.g., Prim's, Kruskal's ;
- The obtained Markov tree T is not a causation relation, (causal models are more difficult to obtain),

Part IV Probabilistic Models and Learning

We note that:

- Finding the best Markov tree equals the maximum spanning tree problem;
- Algorithms for MinST suit MaxST, e.g., Prim's, Kruskal's ;
- The obtained Markov tree T is not a causation relation, (causal models are more difficult to obtain),
- The optimization idea based on D_{KL} has yet to be used to obtain Markov graphs of topologies beyond tree

Part IV Probabilistic Models and Learning

We note that:

- Finding the best Markov tree equals the maximum spanning tree problem;
- Algorithms for MinST suit MaxST, e.g., Prim's, Kruskal's ;
- The obtained Markov tree T is not a causation relation, (causal models are more difficult to obtain),
- The optimization idea based on D_{KL} has yet to be used to obtain Markov graphs of topologies beyond tree until now.