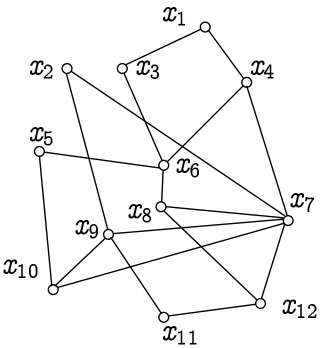
Stat 375: Inference in Graphical Models
Lectures 3-4

Andrea Montanari

Stanford University

March 31, 2012

Undirected Pairwise Graphical Model



$G = (V, E)$, $V = [n]$, $x = (x_1, \dots, x_n)$, $x_i \in \mathcal{X}$, $|\mathcal{X}| < \infty$ (small)

$$\mu(x) = \frac{1}{Z} \prod_{(ij) \in E} \psi_{ij}(x_i, x_j) \,.$$

$$\psi \equiv (\psi_{ij})_{(i,j) \in E}$$

Probabilistic inference tasks

Given (G, ψ)

- Compute the partition function $Z_G(\theta)$
- Compute expectations $\mathbb{E}_\mu\{f(x_i)\}$
- Sample from $\mu_{G,\psi}(\,\cdot\,)$

The above are roughly equivalent.

These two lectures

- Computatonal hardness and reductions.

- Graph structure and hardness: the case of one-dimensional models.

Computational hardness and reductions

Formalization: The class #P

$$
\begin{aligned}
\mathsf{S} &= \{0,1\}^* \equiv \{\text{ binary strings }\}\,,\\
|s| &= \text{length of string } s\,.
\end{aligned}
$$

Formalization: The class #P

A p-relation:

$$R : \mathsf{S} \times \mathsf{S} \to \{0,1\}\,,$$

- $R(s,x) = 1 \;\Rightarrow\; |x| \le \mathrm{Poly}(|s|)$.
- $R(s,x) = 1$ can be checked in polynomial time

Formalization: The class #P

Decision: Given s is there an x such that $R(s,x) = 1$?

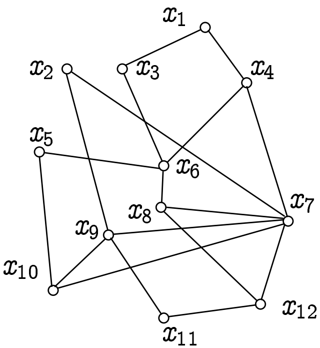
Construction: Given s construct an x such that $R(s,x) = 1$.

Counting: Given s how many x's are such that $R(s,x) = 1$?
(Class $\#P$)

Sampling: Given s, sample a uniformy random element in $\{x : R(s,x) = 1\}$.

Marginals: Need a bit more structure.

Example: Independent sets



$s = G = (V, E)$
$x = (x_i)_{i \in V}$, $x_i \in \mathcal{X} = \{0, 1\}$

$$R_{\text{IS}}(G, x) = \begin{cases} 0 & \text{if there exists } (i,j) \in E \text{ with } x_i = x_j = 1 \\ 1 & \text{otherwise.} \end{cases}$$

Associated graphical model

$$\mu(x) = \frac{1}{Z(G)} \prod_{(i,j)\in E} \psi(x_i, x_j)\,, \qquad \psi(x_i, x_j) = \begin{cases} 0 & \text{if } x_i = x_j = 1, \\ 1 & \text{otherwise.} \end{cases}$$

Counting is computing a partition function.

$$\begin{aligned} Z(G) &= \left|\text{Independent Sets}(G)\right| \\ &= \left|\{\, x \,:\, R_{\rm IS}(G,x) = 1 \,\}\right| \end{aligned}$$

Associated graphical model

$$\mu(x) = \frac{1}{Z(G)} \prod_{(i,j)\in E} \psi(x_i, x_j)\,, \qquad \psi(x_i, x_j) = \begin{cases} 0 & \text{if } x_i = x_j = 1, \\ 1 & \text{otherwise.} \end{cases}$$

Counting is computing a partition function.

$$\begin{aligned} Z(G) &= \big|\text{Independent Sets}(G)\big| \\ &= \big|\{\, x \,:\, R_{\rm IS}(G,x) = 1\,\}\big| \end{aligned}$$

Does #P require $\psi_{ij}(x_i, x_j) \in \{0, 1\}$?

Can have $\psi_{ij} : \mathcal{X} \times \mathcal{X} \to \mathbb{N}$:

- Add factor node and auxiliary variable for edge (i, j).

Can have $\psi_{ij} : \mathcal{X} \times \mathcal{X} \to \mathbb{Q}$:

- Normalize to make all weights integer.

Does #P require $\psi_{ij}(x_i, x_j) \in \{0, 1\}$?

Can have $\psi_{ij} : \mathcal{X} \times \mathcal{X} \to \mathbb{N}$:

- Add factor node and auxiliary variable for edge (i, j).

Can have $\psi_{ij} : \mathcal{X} \times \mathcal{X} \to \mathbb{Q}$:

- Normalize to make all weights integer.

Reductions within #P

Counting R_1 reduces to counting R_2 if you have

- A poly-mapping $\varphi : \mathsf{S} \to \mathsf{S}$
- A poly-mapping $\gamma : \mathbb{N} \to \mathbb{N}$, s.t.

defining

$$Z_{1/2}(s) \equiv \left|\{ x : R_{1/2}(s,x) = 1 \}\right| ,$$

we have

$$Z_1(s) = \gamma(Z_2(\varphi(s)))$$

(Usually one takes $\gamma(N) = N$.)

Reductions within #P

Counting R_1 reduces to counting R_2 if you have

- A poly-mapping $\varphi : \mathsf{S} \to \mathsf{S}$
- A poly-mapping $\gamma : \mathbb{N} \to \mathbb{N}$, s.t.

defining

$$Z_{1/2}(s) \equiv \left|\{ x : \; R_{1/2}(s,x) = 1 \}\right| ,$$

we have

$$Z_1(s) = \gamma(Z_2(\varphi(s)))$$

(Usually one takes $\gamma(N) = N$.)

Bad news

R is #P-complete if any problem in #P can be reduced to it.

Theorem (Valiant 1979)

- *Independent sets is #P complete.*
- *q-coloring ($q \geq 3$) is #P complete.*
- *#-SAT is #P complete.*
- *Permanent is #P complete.*

Bad news

R is #P-complete if any problem in #P can be reduced to it.

Theorem (Valiant 1979)

- *Independent sets is #P complete.*
- *q-coloring ($q \geq 3$) is #P complete.*
- *#-SAT is #P complete.*
- *Permanent is #P complete.*

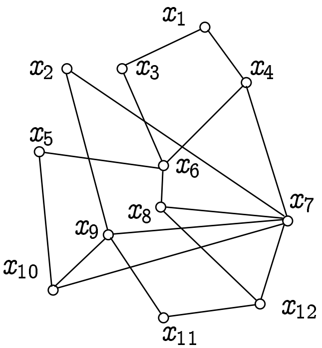
Reductions from other inference problems

Proposition

If R is self-reducible, then sampling and marginals reduce to counting.

Will not define 'self-reducible' formally

Independent sets: Marginals → Counting

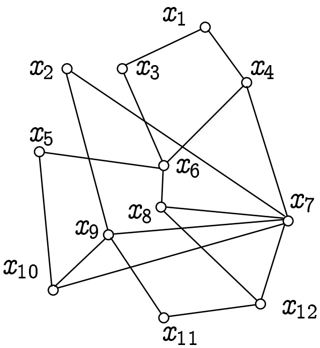


$$\mu(\{x_i = 1\}) = \frac{Z(G, x_i = 1)}{Z(G, x_i = 0) + Z(X_i = 1)},$$

$$Z(G, x_i = \xi) \equiv \left|\{\text{ independent sets of } G \text{ with } x_i = \xi\}\right|,$$

$$Z(G, x_i = 0) = Z(G \setminus \{i\}), \qquad Z(G, x_i = 1) = Z(G \setminus \{i\} \cup \partial i).$$

Independent sets: Marginals → Counting

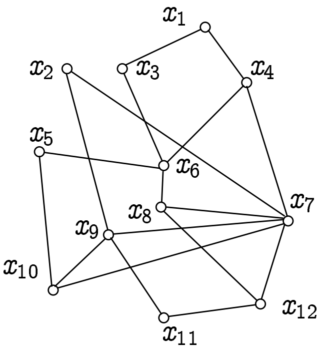


$$\mu(\{x_i = 1\}) = \frac{Z(G, x_i = 1)}{Z(G, x_i = 0) + Z(X_i = 1)},$$

$$Z(G, x_i = \xi) \equiv \left|\{\text{ independent sets of } G \text{ with } x_i = \xi\,\}\right|,$$

$Z(G, x_i = 0) = Z(G \setminus \{i\}), \qquad Z(G, x_i = 1) = Z(G \setminus \{i\} \cup \partial i).$

Independent sets: Marginals → Counting



$$\mu(\{x_i = 1\}) = \frac{Z(G, x_i = 1)}{Z(G, x_i = 0) + Z(X_i = 1)},$$

$$Z(G, x_i = \xi) \equiv \big|\{\text{ independent sets of } G \text{ with } x_i = \xi\,\}\big|,$$

$$Z(G, x_i = 0) = Z(G \setminus \{i\}), \qquad Z(G, x_i = 1) = Z(G \setminus \{i\} \cup \partial i).$$

Independent sets: Sampling $\to$ Marginals

SEQUENTIAL SAMPLING(graph G)

```
Set $G_0 = G$, $\ell = 0$;
While $G_\ell$ is not empty:
    Choose $i \in V(G_\ell)$;
    Compute $p_i = \mu_{G_\ell}(\{x_i = 1\})$;
    Draw $x_i \sim$Bernoulli$(p_i)$;
    If $x_i = 1$, set $G_{\ell+1} = G_\ell \setminus \{i\} \cup \partial i$, $\ell \leftarrow \ell + 1$;
    If $x_i = 0$, set $G_{\ell+1} = G_\ell \setminus \{i\}$, $\ell \leftarrow \ell + 1$;
```

Exercise

How should the above be modified for q-coloring? What about a general MRF?

Independent sets: Sampling $\rightarrow$ Marginals

SEQUENTIAL SAMPLING(graph G)

```
Set G_0 = G, ℓ = 0;
While G_ℓ is not empty:
    Choose i ∈ V(G_ℓ);
    Compute p_i = μ_{G_ℓ}({x_i = 1});
    Draw x_i ~Bernoulli(p_i);
    If x_i = 1, set G_{ℓ+1} = G_ℓ \ {i} ∪ ∂i, ℓ ← ℓ + 1;
    If x_i = 0, set G_{ℓ+1} = G_ℓ \ {i}, ℓ ← ℓ + 1;
```

Exercise

How should the above be modified for q-coloring? What about a general MRF?

Independent sets: Counting $\to$ Marginals

$$\frac{1}{Z(G)} = \frac{Z(G\setminus\{1\})}{Z(G)}\cdot\frac{Z(G\setminus\{1,2\})}{Z(G\setminus\{1\})}\cdots\frac{Z(\emptyset)}{Z(G\setminus\{1,\dots,n-1\})}$$

$$= \mu_G(x_1=0)\cdot\mu_{G\setminus\{1\}}(x_2=0)\cdots\mu_{G\setminus\{1,\dots,n-1\}}(x_n=0)$$

Independent sets: Counting $\rightarrow$ Marginals

$$\begin{aligned}\frac{1}{Z(G)} &= \frac{Z(G\setminus\{1\})}{Z(G)}\cdot\frac{Z(G\setminus\{1,2\})}{Z(G\setminus\{1\})}\cdots\frac{Z(\emptyset)}{Z(G\setminus\{1,\dots,n-1\})}\\ &= \mu_G(x_1=0)\cdot\mu_{G\setminus\{1\}}(x_2=0)\cdots\mu_{G\setminus\{1,\dots,n-1\}}(x_n=0)\end{aligned}$$

Independent sets: Marginals → Sampling

Samples: $x^{(1)}$, $x^{(2)}$,... $x^{(m)} \sim_{i.i.d} \mu_G$

$$\widehat{\mu}_G(x_i = 1) = \frac{1}{m}\sum_{\ell=1}^{m} x_i^{(\ell)}$$

Taking $m = \text{Poly}(n)$, can estimate marginals with relative error $1/\text{Poly}(n)$.

Not exact reduction!

Indeed there are problems for which sampling is easy and exact counting #P hard (e.g. #DNF).

Independent sets: Marginals → Sampling

Samples: $x^{(1)}, x^{(2)}, \dots x^{(m)} \sim_{i.i.d} \mu_G$

$$\widehat{\mu}_G(x_i = 1) = \frac{1}{m} \sum_{\ell=1}^{m} x_i^{(\ell)}$$

Taking $m = \text{Poly}(n)$, can estimate marginals with relative error $1/\text{Poly}(n)$.

Not exact reduction!

Indeed there are problems for which sampling is easy and exact counting #P hard (e.g. #DNF).

Independent sets: Marginals $\to$ Sampling

Samples: $x^{(1)}, x^{(2)}, \dots x^{(m)} \sim_{i.i.d} \mu_G$

$$\widehat{\mu}_G(x_i = 1) = \frac{1}{m} \sum_{\ell=1}^{m} x_i^{(\ell)}$$

Taking $m = \text{Poly}(n)$, can estimate marginals with relative error $1/\text{Poly}(n)$.

Not exact reduction!
Indeed there are problems for which sampling is easy and exact counting #P hard (e.g. #DNF).

The next best thing

FPTAS (Fully Polynomial Time Approximation Scheme)
Given G, ε, outputs $\widehat{Z}(G, \varepsilon)$ in time $\text{Poly}(n, 1/\varepsilon)$ such that

$$(1-\varepsilon)Z(G) \le \widehat{Z}(G,\varepsilon) \le (1+\varepsilon)Z(G)\,.$$

FPRAS (Fully Polynomial time Randomized Approximation Scheme)
Given G, ε, outputs $\widehat{Z}(G, \varepsilon)$ in time $\text{Poly}(n, 1/\varepsilon)$ such that

$$\mathbb{P}\{(1-\varepsilon)Z(G) \le \widehat{Z}(G,\varepsilon) \le (1+\varepsilon)Z(G)\} \ge \frac{3}{4}\,.$$

The next best thing

FPTAS (Fully Polynomial Time Approximation Scheme)
Given G, ε, outputs $\widehat{Z}(G, \varepsilon)$ in time $\text{Poly}(n, 1/\varepsilon)$ such that

$$(1-\varepsilon)Z(G) \le \widehat{Z}(G, \varepsilon) \le (1+\varepsilon)Z(G) .$$

FPRAS (Fully Polynomial time Randomized Approximation Scheme)
Given G, ε, outputs $\widehat{Z}(G, \varepsilon)$ in time $\text{Poly}(n, 1/\varepsilon)$ such that

$$\mathbb{P}\{(1-\varepsilon)Z(G) \le \widehat{Z}(G, \varepsilon) \le (1+\varepsilon)Z(G)\} \ge \frac{3}{4} .$$

Reductions between approximation problems

Counting: Computing partition function within multiplicative error ε.

Marginals: Computing marginals within multiplicative error ε.

Sampling: Generating $x \sim \widetilde{\mu}$, with $||\widetilde{\mu} - \mu||_{TV} \le \varepsilon$.

These are polynomially equivalent.

Reductions between approximation problems

Counting: Computing partition function within multiplicative error ε.

Marginals: Computing marginals within multiplicative error ε.

Sampling: Generating $x \sim \widetilde{\mu}$, with $||\widetilde{\mu} - \mu||_{TV} \leq \varepsilon$.

These are polynomially equivalent.

Bad news and good news

Theorem

No FPRAS exists for { INDEPENDENT SETS, #SAT, ISING, . . . } *unless P=NP.*

[e.g. Luby, Vigoda, 1999,. . . ,Sly 2010]

Theorem

FPRAS/FPTAS exist for { INDEPENDENT SETS AT SMALL DEGREE, PERMANENT, FERROMAGNETIC ISING, . . . }.

[We'll see examples of this type]

Bad news and good news

Theorem

No FPRAS exists for { Independent Sets, #SAT, Ising, . . . } *unless P=NP.*

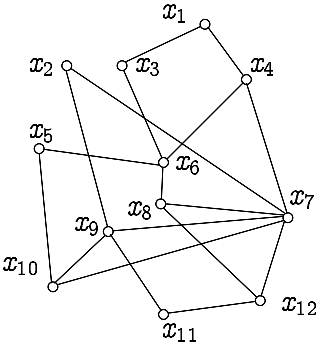
[e.g. Luby, Vigoda, 1999,. . . ,Sly 2010]

Theorem

FPRAS/FPTAS exist for { Independent Sets at small Degree, Permanent, Ferromagnetic Ising, . . . }.

[We'll see examples of this type]

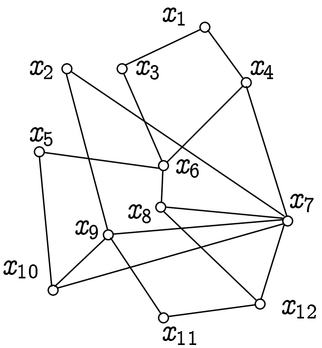
How do you prove inapproximability?



Antiferromagnetic Ising model: $x = (x_1, \dots, x_n) \in \{+1, -1\}^V$

$$\mu(x) = \frac{1}{Z(G,\beta)} \exp\Big\{ -\beta \sum_{(i,j)\in E} x_i x_j \Big\}$$

How do you prove inapproximability?



Antiferromagnetic Ising model: $x = (x_1, \dots, x_n) \in \{+1, -1\}^V$

$$\begin{aligned} \text{cut}_G(x) &\equiv \Big|\Big\{ \text{ edges connecting } x_i = +1 \text{ and } x_j = -1 \Big\}\Big|, \\ \mu(x) &= \frac{1}{Z(G,\beta)} \exp\Big\{2\beta \text{cut}_G(x)\Big\} \end{aligned}$$

For β large enough

$$\mu(x) = \frac{1}{Z(G,\beta)} \exp\left\{2\beta \text{cut}_G(x)\right\}$$

concentrates on MAXCUTS

Exercise

If $\beta = n^2$ then, for all n large enough

$$\mathbb{P}_\mu\{\text{cut}_G(x) = \text{MAXCUT}(G)\} \geq \frac{1}{2}\,.$$

But MAXCUT is NP-complete

For β large enough

$$\mu(x) \quad = \quad \frac{1}{Z(G,\beta)}\exp\left\{2\beta \text{cut}_G(x)\right\}$$

concentrates on MAXCUTS

Exercise

If $\beta = n^2$ then, for all n large enough

$$\mathbb{P}_\mu\{\text{cut}_G(x) = \text{MAXCUT}(G)\} \geq \frac{1}{2}\,.$$

But MAXCUT is NP-complete

For β large enough

$$\mu(x) = \frac{1}{Z(G,\beta)} \exp\left\{2\beta \text{cut}_G(x)\right\}$$

concentrates on MAXCUTS

Exercise

If $\beta = n^2$ then, for all n large enough

$$\mathbb{P}_\mu\{\text{cut}_G(x) = \text{MAXCUT}(G)\} \geq \frac{1}{2}\,.$$

But MAXCUT is NP-complete

But you cheated taking β so large!!!

Keep $\beta = 1$ and

- Replicate each edge n^2 times.

Keep $\beta = 1$ and

- Replicate each vertex k times.
- Place a completely connected graph on the k copies.
- Replace each edge by a bipartite completely connected graph on k vertices.

But you cheated taking β so large!!!

Keep $\beta = 1$ and

- Replicate each edge n^2 times.

Keep $\beta = 1$ and

- Replicate each vertex k times.
- Place a completely connected graph on the k copies.
- Replace each edge by a bipartite completely connected graph on k vertices.

But you cheated taking β so large!!!

Keep $\beta = 1$ and

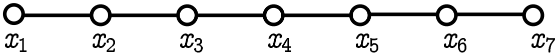
- Replicate each edge n^2 times.

Keep $\beta = 1$ and

- Replicate each vertex k times.
- Place a completely connected graph on the k copies.
- Replace each edge by a bipartite completely connected graph on k vertices.

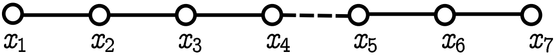
The case of one-dimensional models

Line graph



$$\mu(x) = \frac{1}{Z} \prod_{i=1}^{n-1} \psi_i(x_i, x_{i+1})$$

Line graph: Computing marginals



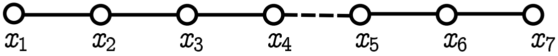
$$\mu_{i\to}(x_1,\dots,x_i) \equiv \frac{1}{Z_{i\to}}\prod_{j=1}^{i-1}\psi_j(x_j,x_{j+1}),$$

$$\mu_{(i+1)\leftarrow}(x_{i+1},\dots,x_n) \equiv \frac{1}{Z_{(i+1)\leftarrow}}\prod_{j=i+1}^{n-1}\psi_j(x_j,x_{j+1}),$$

$$\nu_{i\to}(x_i) \equiv \sum_{x_1,\dots,x_{i-1}}\mu_{i\to}(x_1,\dots,x_i),$$

$$\nu_{(i+1)\leftarrow}(x_{i+1}) \equiv \sum_{x_{i+2},\dots,x_n}\mu_{(i+1)\leftarrow}(x_{i+1},\dots,x_n).$$

Line graph: Computing marginals



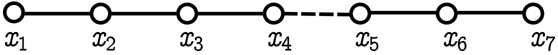
$$\mu_{i\to}(x_1,\dots,x_i) \equiv \frac{1}{Z_{i\to}}\prod_{j=1}^{i-1}\psi_j(x_j,x_{j+1})\,,$$

$$\mu_{(i+1)\leftarrow}(x_{i+1},\dots,x_n) \equiv \frac{1}{Z_{(i+1)\leftarrow}}\prod_{j=i+1}^{n-1}\psi_j(x_j,x_{j+1})\,,$$

$$\nu_{i\to}(x_i) \equiv \sum_{x_1,\dots,x_{i-1}}\mu_{i\to}(x_1,\dots,x_i)\,,$$

$$\nu_{(i+1)\leftarrow}(x_{i+1}) \equiv \sum_{x_{i+2},\dots,x_n}\mu_{(i+1)\leftarrow}(x_{i+1},\dots,x_n)\,.$$

Line graph: Computing marginals

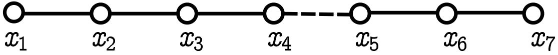


$$\mu_{i\to}(x_1,\dots,x_i) = \frac{1}{\widetilde{Z}_{i\to}}\,\mu_{(i-1)\to}(x_1,\dots,x_{i-1})\,\psi_{i-1}(x_{i-1},x_i)\,,$$

$$\sum_{x_1,\dots,x_{i-1}} \mu_{i\to}(x_1,\dots,x_i) = \frac{1}{\widetilde{Z}_{i\to}} \sum_{x_1,\dots,x_{i-1}} \mu_{(i-1)\to}(x_1,\dots,x_{i-1})\,\psi_{i-1}(x_{i-1},x_i)\,,$$

$$\nu_{i\to}(x_i) = \frac{1}{\widetilde{Z}_{i\to}} \sum_{x_{i-1}} \nu_{(i-1)\to}(x_{i-1})\psi_{i-1}(x_{i-1},x_i)\,,$$

Line graph: Computing marginals

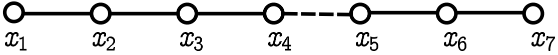


$$\mu_{i\to}(x_1,\dots,x_i) = \frac{1}{\widetilde{Z}_{i\to}}\,\mu_{(i-1)\to}(x_1,\dots,x_{i-1})\,\psi_{i-1}(x_{i-1},x_i)\,,$$

$$\sum_{x_1,\dots,x_{i-1}}\mu_{i\to}(x_1,\dots,x_i) = \frac{1}{\widetilde{Z}_{i\to}}\sum_{x_1,\dots,x_{i-1}}\mu_{(i-1)\to}(x_1,\dots,x_{i-1})\,\psi_{i-1}(x_{i-1},x_i)\,,$$

$$\nu_{i\to}(x_i) = \frac{1}{\widetilde{Z}_{i\to}}\sum_{x_{i-1}}\nu_{(i-1)\to}(x_{i-1})\psi_{i-1}(x_{i-1},x_i)\,,$$

Line graph: Computing marginals

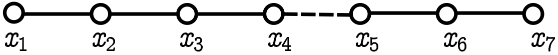


$$\mu_{i\to}(x_1,\dots,x_i) = \frac{1}{\widetilde{Z}_{i\to}}\,\mu_{(i-1)\to}(x_1,\dots,x_{i-1})\,\psi_{i-1}(x_{i-1},x_i)\,,$$

$$\sum_{x_1,\dots,x_{i-1}}\mu_{i\to}(x_1,\dots,x_i) = \frac{1}{\widetilde{Z}_{i\to}}\sum_{x_1,\dots,x_{i-1}}\mu_{(i-1)\to}(x_1,\dots,x_{i-1})\,\psi_{i-1}(x_{i-1},x_i)\,,$$

$$\nu_{i\to}(x_i) = \frac{1}{\widetilde{Z}_{i\to}}\sum_{x_{i-1}}\nu_{(i-1)\to}(x_{i-1})\psi_{i-1}(x_{i-1},x_i)\,,$$

Line graph: Computing marginals

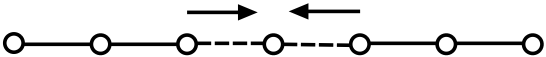


$$\mu_{i\to}(x_1,\dots,x_i) \cong \mu_{(i-1)\to}(x_1,\dots,x_{i-1})\,\psi_{i-1}(x_{i-1},x_i)\,,$$

$$\sum_{x_1,\dots,x_{i-1}} \mu_{i\to}(x_1,\dots,x_i) \cong \sum_{x_1,\dots,x_{i-1}} \mu_{(i-1)\to}(x_1,\dots,x_{i-1})\,\psi_{i-1}(x_{i-1},x_i)\,,$$

$$\nu_{i\to}(x_i) \cong \sum_{x_{i-1}} \nu_{(i-1)\to}(x_{i-1})\,\psi_{i-1}(x_{i-1},x_i)\,,$$

Line graph: Computing marginals

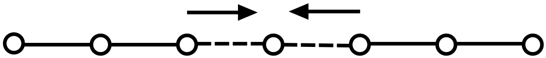


$$\nu_{i\to}(x_i) \cong \sum_{x_{i-1}} \nu_{(i-1)\to}(x_{i-1})\, \psi_{i-1}(x_{i-1}, x_i)\,,$$

$$\nu_{i\leftarrow}(x_i) \cong \sum_{x_{i+1}} \nu_{(i+1)\leftarrow}(x_{i+1})\, \psi_{i}(x_{i}, x_{i+1})\,,$$

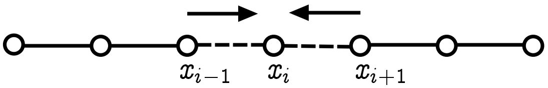
$$\nu_{i}(x_i) \cong \sum_{x_{i-1}, x_{i+1}} \psi_{i-1}(x_{i-1}, x_i)\psi_{i}(x_{i}, x_{i+1})\, \nu_{(i-1)\to}(x_{i-1})\, \nu_{(i+1)\leftarrow}(x_{i+1})\,.$$

Line graph: Computing marginals



$$
\begin{aligned}
\nu_{i\to}(x_i) &\cong \sum_{x_{i-1}} \nu_{(i-1)\to}(x_{i-1})\,\psi_{i-1}(x_{i-1},x_i)\,, \\
\nu_{i\leftarrow}(x_i) &\cong \sum_{x_{i+1}} \nu_{(i+1)\leftarrow}(x_{i+1})\,\psi_i(x_i,x_{i+1})\,, \\
\nu_i(x_i) &\cong \sum_{x_{i-1},x_{i+1}} \psi_{i-1}(x_{i-1},x_i)\psi_i(x_i,x_{i+1})\,\nu_{(i-1)\to}(x_{i-1})\,\nu_{(i+1)\leftarrow}(x_{i+1})\;.
\end{aligned}
$$

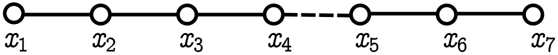
Line graph: Computing marginals



$$\mu(x_i) = \nu_i(x_i)$$

Marginals can be computed in $|\mathcal{X}|^2 n$ operations.

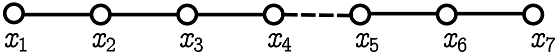
Line graph: Computing the partition function



$$
\begin{aligned}
Z_{i\to} &= Z_{(i-1)\to}\widetilde{Z}_{i\to}\,, \\
\widetilde{Z}_{i\to} &= \sum_{x_i,x_{i-1}} \nu_{(i-1)\to}(x_{i-1})\,\psi_{i-1}(x_{i-1},x_i)\,, \\
Z_{1\to} &= 1\,, \\
Z_{n\to} &= Z\,.
\end{aligned}
$$

The partition function can be computed in $|\mathcal{X}|^2 n$ operations.

Line graph: Computing the partition function



$$
\begin{aligned}
Z_{i\to} &= Z_{(i-1)\to}\widetilde{Z}_{i\to}\,, \\
\widetilde{Z}_{i\to} &= \sum_{x_i,x_{i-1}} \nu_{(i-1)\to}(x_{i-1})\,\psi_{i-1}(x_{i-1},x_i)\,, \\
Z_{1\to} &= 1\,, \\
Z_{n\to} &= Z\,.
\end{aligned}
$$

The partition function can be computed in $|\mathcal{X}|^2 n$ operations.

Example #1: Hidden Markov Models

Sequence of r.v.'s $\{(X_1, Y_1); (X_2, Y_2); \ldots ; (X_n, Y_n)\}$,

$\{X_i\}$ Markov Chain $\mathbb{P}\{x\} = \mathbb{P}\{x_1\} \prod_{i=1}^{n-1} \mathbb{P}\{x_{i+1}|x_i\}$,

$\{Y_i\}$ noisy observations $\mathbb{P}\{y|x\} = \prod_{i=i}^{n} \mathbb{P}\{y_i|x_i\}$,

Example #1: Hidden Markov Models

Sequence of r.v.'s $\{(X_1, Y_1); (X_2, Y_2); \dots; (X_n, Y_n)\}$,

$\{X_i\}$ Markov Chain $\mathbb{P}\{x\} = \mathbb{P}\{x_1\} \prod_{i=1}^{n-1} \mathbb{P}\{x_{i+1}|x_i\}$,

$\{Y_i\}$ noisy observations $\mathbb{P}\{y|x\} = \prod_{i=i}^{n} \mathbb{P}\{y_i|x_i\}$,

Example #1: Hidden Markov Models

Sequence of r.v.'s $\{(X_1, Y_1); (X_2, Y_2); \dots ; (X_n, Y_n)\}$,

$\{X_i\}$ Markov Chain $\mathbb{P}\{x\} = \mathbb{P}\{x_1\} \prod_{i=1}^{n-1} \mathbb{P}\{x_{i+1}|x_i\}$,

$\{Y_i\}$ noisy observations $\mathbb{P}\{y|x\} = \prod_{i=i}^{n} \mathbb{P}\{y_i|x_i\}$,

Example #1: Time Homogeneous Hidden Markov Models

Sequence of r.v.'s $\quad \{(X_1, Y_1); (X_2, Y_2); \dots; (X_n, Y_n)\}\,,$

$\{X_i\}$ Markov Chain $\quad \mathbb{P}\{x\} = q_0(x_1) \prod_{i=1}^{n-1} q(x_i, x_{i+1})\,,$

$\{Y_i\}$ noisy observations $\quad \mathbb{P}\{y|x\} = \prod_{i=i}^{n} r(x_i, y_i)\,,$

Parameters: $\{p(x, x') : \; x, x' \in \mathcal{X}\}, \{r(x, y) : \; x \in \mathcal{X},\, y \in \mathcal{Y}\}$

Example #1:
Time Homogeneous Hidden Markov Models

Sequence of r.v.'s $\{(X_1, Y_1); (X_2, Y_2); \dots; (X_n, Y_n)\}$,

$\{X_i\}$ Markov Chain $\mathbb{P}\{x\} = q_0(x_1) \prod_{i=1}^{n-1} q(x_i, x_{i+1})$,

$\{Y_i\}$ noisy observations $\mathbb{P}\{y|x\} = \prod_{i=i}^{n} r(x_i, y_i)$,

Parameters: $\{p(x, x') : x, x' \in \mathcal{X}\}, \{r(x, y) : x \in \mathcal{X}, y \in \mathcal{Y}\}$

Example #1: Time Homogeneous Hidden Markov Models

Sequence of r.v.'s $\{(X_1, Y_1); (X_2, Y_2); \ldots; (X_n, Y_n)\}$,

$\{X_i\}$ Markov Chain $\mathbb{P}\{x\} = q_0(x_1) \prod_{i=1}^{n-1} q(x_i, x_{i+1})$,

$\{Y_i\}$ noisy observations $\mathbb{P}\{y|x\} = \prod_{i=i}^{n} r(x_i, y_i)$,

Parameters: $\{p(x, x') : x, x' \in \mathcal{X}\}, \{r(x, y) : x \in \mathcal{X}, y \in \mathcal{Y}\}$

Example #1: Time Homogeneous Hidden Markov Models

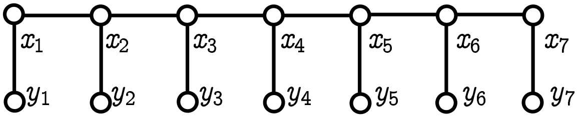
Sequence of r.v.'s $\{(X_1, Y_1); (X_2, Y_2); \dots; (X_n, Y_n)\}$,

$\{X_i\}$ Markov Chain $\mathbb{P}\{x\} = q_0(x_1) \prod_{i=1}^{n-1} q(x_i, x_{i+1})$,

$\{Y_i\}$ noisy observations $\mathbb{P}\{y|x\} = \prod_{i=i}^{n} r(x_i, y_i)$,

Parameters: $\{p(x, x') : x, x' \in \mathcal{X}\}, \{r(x, y) : x \in \mathcal{X}, y \in \mathcal{Y}\}$

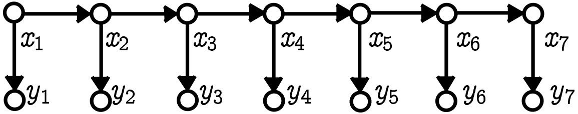
Example #1: Unconditional
Time Homogeneous Hidden Markov Models



$$\mu(x,y) = \frac{1}{Z}\prod_{i=1}^{n-1}\psi_i(x_i,x_{i+1})\prod_{i=1}^{n}\tilde{\psi}_i(x_i,y_i)\,,$$

$$\psi_i(x_i,x_{i+1}) = q(x_i,x_{i+1})\,,\qquad \tilde{\psi}_i(x_i,y_i) = r(x_i,y_i)\,.$$

(equivalent directed form



Question

Estimate $\widehat{x}(y)$ from y as to minimize

$$\mathbb{E}N_{\rm err} = \sum_{i=1}^{n} \mathbb{P}\{\widehat{x}_i(y) \neq x_i\}\,.$$

Exercise: The optimal estimator is

$$\widehat{x}_i(y) = \arg\max_{z\in\mathcal{X}} \mathbb{P}\{x_i = z|y\}\,.$$

Computational task.

Compute marginals of (q_0 uniform)

$$\mu_y(x) = \mathbb{P}\{x|y\} \stackrel{\text{Bayes thm}}{=} \frac{1}{Z(y)} \prod_{i=1}^{n-1} q(x_i, x_{i+1}) \prod_{i=1}^{n} r(x_i, y_i)\,.$$

Question

Estimate $\widehat{x}(y)$ from y as to minimize

$$\mathbb{E}N_{\rm err} = \sum_{i=1}^{n} \mathbb{P}\{\widehat{x}_i(y) \neq x_i\}\,.$$

Exercise: The optimal estimator is

$$\widehat{x}_i(y) = \arg\max_{z\in\mathcal{X}} \mathbb{P}\{x_i = z|y\}\,.$$

Computational task.

Compute marginals of (q_0 uniform)

$$\mu_y(x) = \mathbb{P}\{x|y\} \stackrel{\text{Bayes thm}}{=} \frac{1}{Z(y)} \prod_{i=1}^{n-1} q(x_i, x_{i+1}) \prod_{i=1}^{n} r(x_i, y_i)\,.$$

Question

Estimate $\widehat{x}(y)$ from y as to minimize

$$\mathbb{E}N_{\rm err} = \sum_{i=1}^{n} \mathbb{P}\{\widehat{x}_i(y) \neq x_i\}\,.$$

Exercise: The optimal estimator is

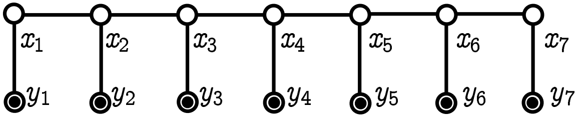
$$\widehat{x}_i(y) = \arg\max_{z\in\mathcal{X}} \mathbb{P}\{x_i = z|y\}\,.$$

Computational task.

Compute marginals of (q_0 uniform)

$$\mu_y(x) = \mathbb{P}\{x|y\} \overset{\text{Bayes thm}}{=} \frac{1}{Z(y)}\prod_{i=1}^{n-1} q(x_i, x_{i+1}) \prod_{i=1}^{n} r(x_i, y_i)\,.$$

Example #1: Conditional Hidden Markov Models

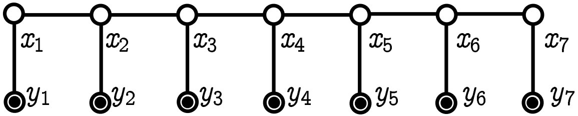


$$\begin{aligned}
\mu_y(x) &= \frac{1}{Z(y)} \prod_{i=1}^{n-1} q(x_i, x_{i+1}) \prod_{i=1}^{n} r(x_i, y_i) \\
&= \frac{1}{Z(y)} \prod_{i=1}^{n-1} \psi_i(x_i, x_{i+1})
\end{aligned}$$

$$\psi_i(x_i, x_{i+1}) = q(x_i, x_{i+1})\, r(x_i, y_i) \qquad (\text{for } i < n-1)$$

$$\psi_{n-1}(x_{n-1}, x_n) = q(x_{n-1}, x_n)\, r(x_{n-1}, y_{n-1})\, r(x_n, y_n)\,.$$

Example #1: Conditional Hidden Markov Models



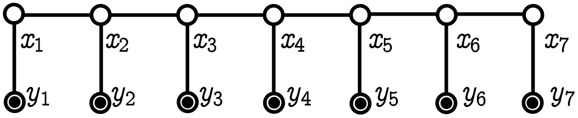
$$\mu_y(x) = \frac{1}{Z(y)} \prod_{i=1}^{n-1} q(x_i, x_{i+1}) \prod_{i=1}^{n} r(x_i, y_i)$$

$$= \frac{1}{Z(y)} \prod_{i=1}^{n-1} \psi_i(x_i, x_{i+1})$$

$$\psi_i(x_i, x_{i+1}) = q(x_i, x_{i+1})\, r(x_i, y_i) \qquad (\text{for } i < n-1)$$

$$\psi_{n-1}(x_{n-1}, x_n) = q(x_{n-1}, x_n)\, r(x_{n-1}, y_{n-1})\, r(x_n, y_n)\,.$$

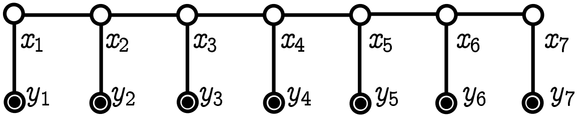
Example #1: Computing marginals in HMMs



$$
\begin{aligned}
\nu_{i\to}(x_i) &\cong \sum_{x_{i-1}\in\mathcal{X}} q(x_{i-1},x_i)\, r(x_{i-1},y_{i-1})\, \nu_{(i-1)\to}(x_{i-1})\,,\\
\nu_{i\leftarrow}(x_i) &\cong \sum_{x_{i+1}\in\mathcal{X}} q(x_i,x_{i+1})\, r(x_i,y_i)\, \nu_{(i+1)\leftarrow}(x_{i+1})\,.
\end{aligned}
$$

(forward-backward algorithm)

Example #1: Computing marginals in HMMs



$$\nu_{i\to}(x_i) \cong \sum_{x_{i-1}\in\mathcal{X}} q(x_{i-1},x_i)\, r(x_{i-1},y_{i-1})\, \nu_{(i-1)\to}(x_{i-1})\,,$$

$$\nu_{i\leftarrow}(x_i) \cong \sum_{x_{i+1}\in\mathcal{X}} q(x_i,x_{i+1})\, r(x_i,y_i)\, \nu_{(i+1)\leftarrow}(x_{i+1})\,.$$

(forward-backward algorithm)

Example #1B: Linear system

HMM with $x_i \in \mathbb{R}^k$, $y_i \in \mathbb{R}^h$

Markov chain (parameters $F, Q \in \mathbb{R}^{k\times k}$, $Q \succeq 0$)

$$
\begin{aligned}
q(x_i, x_{i+1}) &\cong \exp\Big\{ -\frac{1}{2}(x_{i+1} - Fx_i)^T Q^{-1}(x_{i+1} - Fx_i)\Big\}\,, \\
x_i &\sim \mathrm{N}(Fx_i, Q)\,.
\end{aligned}
$$

Observations (parameters $H \in \mathbb{R}^{h\times k}$, $R \in \mathbb{R}^{h\times h}$, $R \succeq 0$)

$$
\begin{aligned}
r(x_i, y_i) &\cong \exp\Big\{ -\frac{1}{2}(y_i - Hx_i)^T R^{-1}(y_i - Hx_i)\Big\}\,, \\
y_i &\sim \mathrm{N}(Hx_i, R)\,.
\end{aligned}
$$

Forward iteration = Kalman filter

$$
\nu_{i\to}(x_i) \cong \exp\Big\{ -\frac{1}{2}(x_i - \widehat{x}_i)^T P_i^{-1}(x_i - \widehat{x}_i)\Big\}
$$

Example #1B: Linear system

HMM with $x_i \in \mathbb{R}^k$, $y_i \in \mathbb{R}^h$
Markov chain (parameters $F, Q \in \mathbb{R}^{k\times k}$, $Q \succeq 0$)

$$
\begin{aligned}
q(x_i, x_{i+1}) &\cong \exp\Big\{ -\frac{1}{2}(x_{i+1} - Fx_i)^T Q^{-1}(x_{i+1} - Fx_i)\Big\}\,,\\
x_i &\sim \mathrm{N}(Fx_i, Q)\,.
\end{aligned}
$$

Observations (parameters $H \in \mathbb{R}^{h\times k}$, $R \in \mathbb{R}^{h\times h}$, $R \succeq 0$)

$$
\begin{aligned}
r(x_i, y_i) &\cong \exp\Big\{ -\frac{1}{2}(y_i - Hx_i)^T R^{-1}(y_i - Hx_i)\Big\}\,,\\
y_i &\sim \mathrm{N}(Hx_i, R)\,.
\end{aligned}
$$

Forward iteration = Kalman filter

$$
\nu_{i\to}(x_i) \cong \exp\Big\{ -\frac{1}{2}(x_i - \widehat{x}_i)^T P_i^{-1}(x_i - \widehat{x}_i)\Big\}
$$

Example #1B: Linear system

HMM with $x_i \in \mathbb{R}^k$, $y_i \in \mathbb{R}^h$
Markov chain (parameters $F, Q \in \mathbb{R}^{k\times k}$, $Q \succeq 0$)

$$\begin{aligned} q(x_i, x_{i+1}) &\cong \exp\Big\{ -\frac{1}{2}(x_{i+1} - Fx_i)^T Q^{-1}(x_{i+1} - Fx_i)\Big\}, \\ x_i &\sim \mathrm{N}(Fx_i, Q)\,. \end{aligned}$$

Observations (parameters $H \in \mathbb{R}^{h\times k}$, $R \in \mathbb{R}^{h\times h}$, $R \succeq 0$)

$$\begin{aligned} r(x_i, y_i) &\cong \exp\Big\{ -\frac{1}{2}(y_i - Hx_i)^T R^{-1}(y_i - Hx_i)\Big\}, \\ y_i &\sim \mathrm{N}(Hx_i, R)\,. \end{aligned}$$

Forward iteration = Kalman filter

$$\nu_{i\to}(x_i) \cong \exp\Big\{ -\frac{1}{2}(x_i - \widehat{x}_i)^T P_i^{-1}(x_i - \widehat{x}_i)\Big\}$$

Example #1C: Neuron firing patterns

Hypothesis

Assemblies of neurones activate in a coordinate way in correspondence to specific cognitive functions. Performing of the function corresponds sequence of these activity states.

Approach

Firing process	$\leftrightarrow$	Observed variables
Activity states	$\leftrightarrow$	Hidden variables

Example #1C: Neuron firing patterns

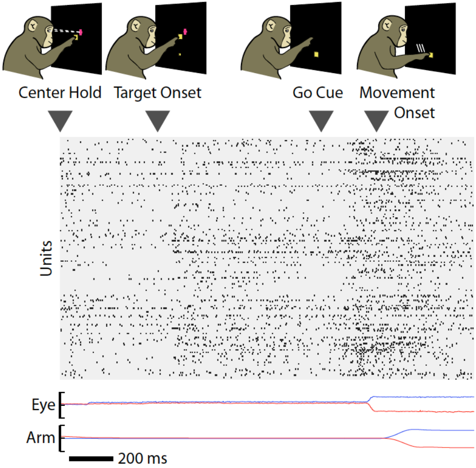
Hypothesis

Assemblies of neurones activate in a coordinate way in correspondence to specific cognitive functions. Performing of the function corresponds sequence of these activity states.

Approach

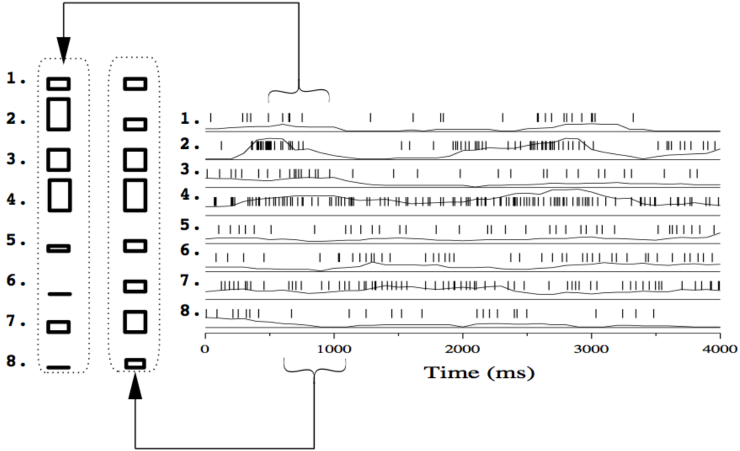
Firing process	$\leftrightarrow$	Observed variables
Activity states	$\leftrightarrow$	Hidden variables

Example #1C: Neuron firing patterns



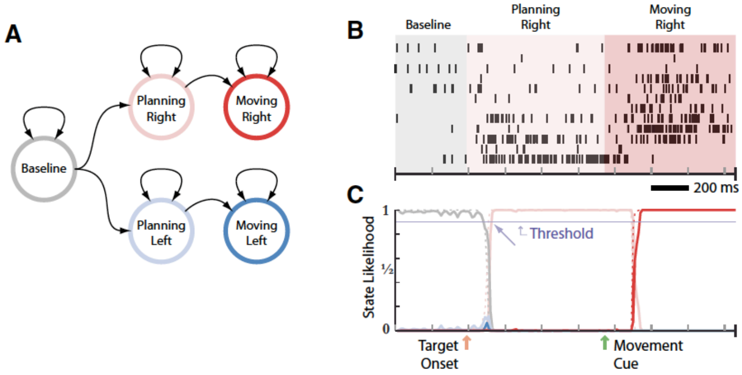
[C. Kemere, G. Santhanam, B. M. Yu, A. Afshar, S.I. Ryu, T. H. Meng and K.V. Shenoy, J. Neurophysiol. 100:2441-2452 (2008)]

Example #1C: Neuron firing patterns



[I. Gat, N. Tishby, and M. Abeles, Network: Computation in Neural Systems, 8:297-322 (1997)]

Example #1C: Neuron firing patterns



[C. Kemere, G. Santhanam, B. M. Yu, A. Afshar, S.I. Ryu, T. H. Meng and K.V. Shenoy, J. Neurophysiol. 100:2441-2452 (2008)]