Stat 375: Inference in Graphical Models
Lectures 1-2

Andrea Montanari

Stanford University

April 2, 2012

Logistics

Mon-Wed 9:30AM-10:45PM, Green Earth Sciences 131

Andrea Montanari (montanari)

http://www.stanford.edu/ montanar/TEACHING/Stat375/stat375.html

Homework due on Mon (starting 4/9/2012).
Take-home final.

Topics

- Probability distributions that are 'local' wrt a graph
- Random variables sit on vertices.
- Strongly dependent if they are nearby.
- *Images, philogenies, error-correcting codes, Bayes networks...*

Topics

Emphasis on computational and mathematical aspects.

Topics

- Equivalent graphical representations.
- Polynomial reductions between various probabilistic inference tasks. Computational hardness.
- Models on trees. Belief propagation.
- Variational inference: naive mean field, Bethe free energy, generalized BP and convex relaxations.
- Gaussian graphical models.
- Learning graphical model form data.
- Correlation decay. The Markov Chain Monte Carlo method.
- Applications to clustering and classification.

Why you should not take this class

- Emphasis on fundamental challenges.
- There is no textbook.
- We'll use the P-word.

Why you should not take this class

- Emphasis on fundamental challenges.
- There is no textbook.
- We'll use the P-word.

Why you should not take this class

- Emphasis on fundamental challenges.
- There is no textbook.
- We'll use the P-word.

Why you should not take this class

- Emphasis on fundamental challenges.
- There is no textbook.
- We'll use the P-word.

Let us start: Families of graphical models

General theme

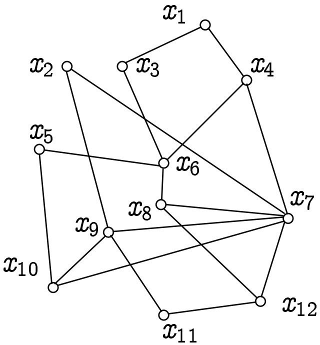
Probability distribution over $x = (x_1, x_2, \dots, x_n)$

$$\mu(x_1, x_2, \dots, x_n)$$

Family # 1: Undirected Pairwise Graphical Models

Family # 1: Undirected Pairwise Graphical Models

(aka Markov Random Fields)



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

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

Undirected Pairwise Graphical Models

Specified by

- Graph $G = (V, E)$.
- Alphabet $\mathcal{X}$.
- Compatibility functions $\psi_{ij} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$, $(i, j) \in E$.

Alphabet

Typically $|\mathcal{X}| < \infty$.

Occasionally $\mathcal{X} = \mathbb{R}$ and

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

(all formulae interpreted as densities)

Key challenge: $n \gg 1$ (space dimension).

Alphabet

Typically $|\mathcal{X}| < \infty$.

Occasionally $\mathcal{X} = \mathbb{R}$ and

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

(all formulae interpreted as densities)

Key challenge: $n \gg 1$ (space dimension).

Alphabet

Typically $|\mathcal{X}| < \infty$.

Occasionally $\mathcal{X} = \mathbb{R}$ and

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

(all formulae interpreted as densities)

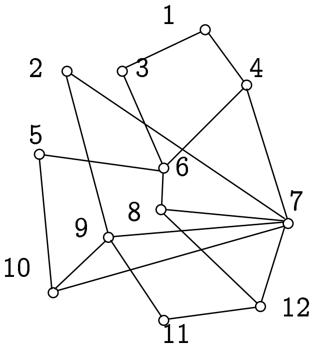
Key challenge: $n \gg 1$ (space dimension).

Partition function

$$Z \equiv \sum_{x \in \mathcal{X}^V} \prod_{(ij) \in E} \psi_{ij}(x_i, x_j)$$

[Plays a crucial role!]

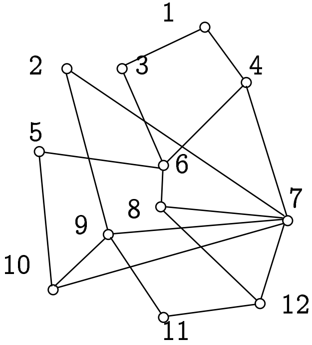
Notations



$$\partial i \equiv \{\text{neighbors of vertex } i\}, \qquad \deg(i) = |\partial i|,$$
$$\partial 9 = \{10, 2, 11\}, \qquad \deg(9) = 3,$$

$$x_U \equiv (x_i)_{i \in U},$$
$$x_{\{1,7,10\}} = (x_1, x_7, x_{10}).$$

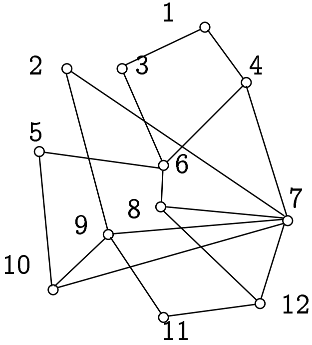
Notations



$$\partial i \equiv \{\text{neighbors of vertex } i\}, \qquad \deg(i) = |\partial i|,$$
$$\partial 9 = \{10, 2, 11\}, \qquad \deg(9) = 3,$$

$$x_U \equiv (x_i)_{i \in U},$$
$$x_{\{1,7,10\}} = (x_1, x_7, x_{10}).$$

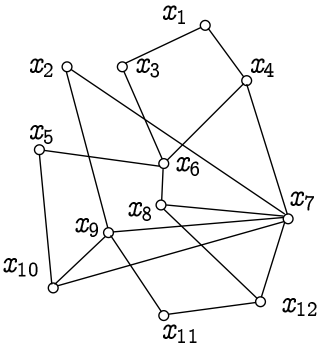
Notations



$$\partial i \equiv \{\text{neighbors of vertex } i\}, \qquad \deg(i) = |\partial i|,$$
$$\partial 9 = \{10, 2, 11\}, \qquad \deg(9) = 3,$$

$$x_U \equiv (x_i)_{i \in U},$$
$$x_{\{1,7,10\}} = (x_1, x_7, x_{10}).$$

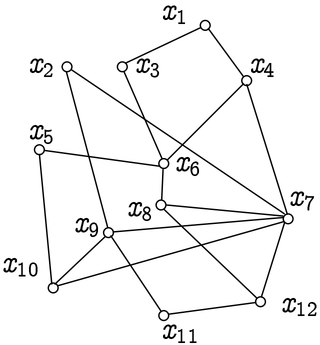
Example: Ising models



$G = (V, E)$, $V = [n]$, $x = (x_1, \dots, x_n)$, $x_i \in \{+1, -1\}$

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

Example: Ising models



$G = (V, E)$, $V = [n]$, $x = (x_1, \dots, x_n)$, $x_i \in \{+1, -1\}$

$$\mu(x) = \frac{1}{Z} \, \exp\Big\{ \sum_{(i,j)\in E} \theta_{ij} x_i x_j \, + \sum_{i \in V} \theta_i x_i \Big\} \, .$$

A motivation: Boltzmann Machines

Can we train a computer to do handwriting?

Can we train a computer to do handwriting?

MNIST dataset: $60,000$ handwritted digits (28×28 pixels)

Can we learn $\mu(x_I)$, $x_I \in \{+1,-1\}^I$, $I = [28] \times [28]$ that generates samples as above?

Can we train a computer to do handwriting?

MNIST dataset: $60,000$ handwritted digits (28×28 pixels)

Can we learn $\mu(x_I)$, $x_I \in \{+1, -1\}^I$, $I = [28] \times [28]$ that generates samples as above?

An attempt

(R. Salakhutdinov, G. Hinton, AISTATS 2009)

What's the magic?

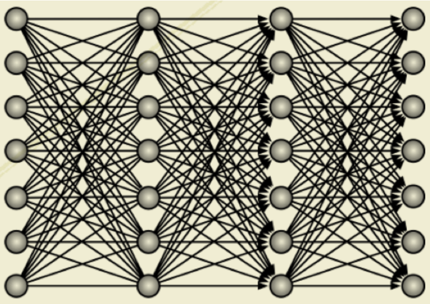
An attempt

(R. Salakhutdinov, G. Hinton, AISTATS 2009)

What's the magic?

What's the magic?

Hidden variables



Image

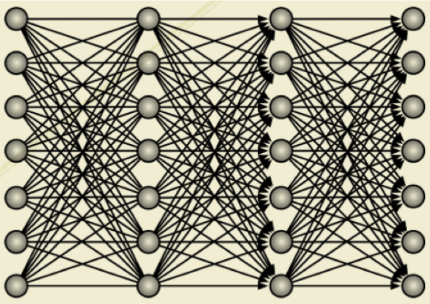
$$\mu(x_I) = \sum_{x_{H(1)},x_{H(2)},x_{H(3)}} \mu_{G,\theta}(x_I, x_{H(1)}, x_{H(2)}, x_{H(3)})$$

$\mu_{G,\theta}(\cdot)$ Ising model on $G = (V, E)$

$$V = (I, H(1), H(2), H(3))$$

What's the magic?

Hidden variables



Image

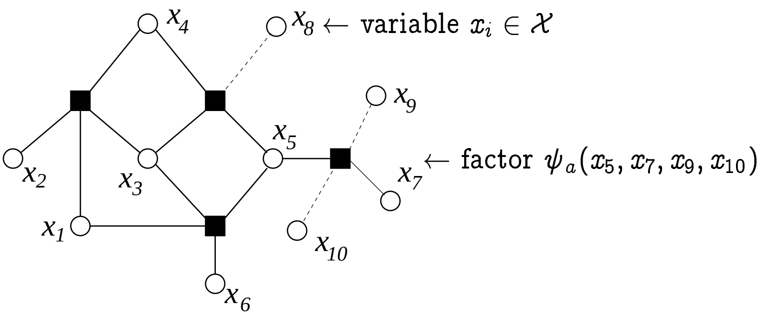
$$\mu(x_I) = \sum_{x_{H(1)}, x_{H(2)}, x_{H(3)}} \mu_{G,\theta}(x_I, x_{H(1)}, x_{H(2)}, x_{H(3)})$$

$\mu_{G,\theta}(\,\cdot\,)$ Ising model on $G = (V, E)$

$$V = (I, H(1), H(2), H(3))$$

Family # 2: Factor Graph Models

Family # 2: Factor Graph Models



$G = (V, F, E)$, $V = [n]$, $x = (x_1, \ldots, x_n)$, $x_i \in \mathcal{X}$

$$\mu(x) = \frac{1}{Z} \prod_{a \in F} \psi_a(x_{\partial a}) .$$

Factor graph models

Terminology

- Variable nodes: $i, j, k, \cdots \in V$.
- Function nodes: $a, b, c \cdots \in F$.

Specified by

- Factor graph $G = (V, F, E)$.
- Alphabet $\mathcal{X}$.
- Compatibility fuctions $\psi_a : \mathcal{X}^{\partial a} \to \mathbb{R}_+$, $a \in F$.

Factor graph models

Terminology

- Variable nodes: $i, j, k, \cdots \in V$.
- Function nodes: $a, b, c \cdots \in F$.

Specified by

- Factor graph $G = (V, F, E)$.
- Alphabet $\mathcal{X}$.
- Compatibility fuctions $\psi_a : \mathcal{X}^{\partial a} \to \mathbb{R}_+$, $a \in F$.

Factor graphs are as powerful as pairwise models

A pairwise model on $G = (V, E)$ with alphabet $\mathcal{X}$ can be represented by a factor graph model on $G' = (V', F', E')$ with $V' = V$, $F' \simeq E$, $|E'| = 2|E|$, $\mathcal{X}' = \mathcal{X}$.

- Put a factor node on each edge.

A factor model on $G = (V, F, E)$ with alphabet $\mathcal{X}$ can be represented by a pairwise model on $G' = (V', F')$ with $V' = V \cup F$, $E' = E$, $\mathcal{X}' = \mathcal{X}^{\Delta}$, $\Delta \equiv \max_{a \in F} \deg(a)$.

- Represent at a factor node the state of its neighbors,

Factor graphs are as powerful as pairwise models

A pairwise model on $G = (V, E)$ with alphabet $\mathcal{X}$ can be represented by a factor graph model on $G' = (V', F', E')$ with $V' = V$, $F' \simeq E$, $|E'| = 2|E|$, $\mathcal{X}' = \mathcal{X}$.

- Put a factor node on each edge.

A factor model on $G = (V, F, E)$ with alphabet $\mathcal{X}$ can be represented by a pairwise model on $G' = (V', F')$ with $V' = V \cup F$, $E' = E$, $\mathcal{X}' = \mathcal{X}^{\Delta}$, $\Delta \equiv \max_{a \in F} \deg(a)$.

- Represent at a factor node the state of its neighbors,

Factor graphs are as powerful as pairwise models

A pairwise model on $G = (V, E)$ with alphabet $\mathcal{X}$ can be represented by a factor graph model on $G' = (V', F', E')$ with $V' = V$, $F' \simeq E$, $|E'| = 2|E|$, $\mathcal{X}' = \mathcal{X}$.

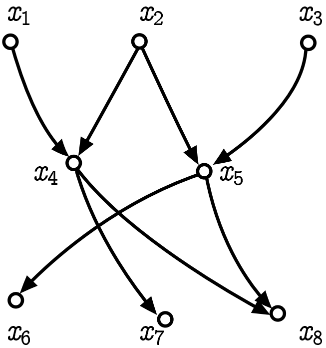
- Put a factor node on each edge.

A factor model on $G = (V, F, E)$ with alphabet $\mathcal{X}$ can be represented by a pairwise model on $G' = (V', F')$ with $V' = V \cup F$, $E' = E$, $\mathcal{X}' = \mathcal{X}^{\Delta}$, $\Delta \equiv \max_{a \in F} \deg(a)$.

- Represent at a factor node the state of its neighbors,

Family # 3: Bayesian Networks

Family # 3: Bayesian Networks



$G = (V, D)$, directed $V = [n]$, $x = (x_1, \dots, x_n)$, $x_i \in \mathcal{X}$
$D = \{$ directed edges $\}$

$$\mu(x) \;=\; \prod_{i \in V} \mu_i(x_i | x_{\pi(i)}) \,, \qquad\qquad \pi(i) = \{ \text{ parents of } i \ \} \,.$$

Bayesian Networks

Specified by

- Directed acyclic graph $G = (V, D)$, $\quad ((i,j) \neq (j,i))$.
- Alphabet $\mathcal{X}$.
- Conditional probability tables $\mu_i(\,\cdot\,|\,\cdot\,) : \mathcal{X} \times \mathcal{X}^{\pi(i)} \to \mathbb{R}_+$, $i \in F$:

$$\sum_{x_i \in \mathcal{X}} \mu_i(x_i|x_{\pi(i)}) = 1 \quad \text{for all } x_{\pi(i)} \in \mathcal{X}^{\pi(i)}\,.$$

Bayes networks are as powerful as factor graphs

A Bayes network $G = (V, D)$ with alphabet $\mathcal{X}$ can be represented by a factor graph model on $G' = (V', F', E')$ with $V' = V$, $|F'| = |V|$, $|E'| = |D| + |V|$, $\mathcal{X}' = \mathcal{X}$.

- Represent by a factor node each CPT.

A factor model on $G = (V, F, E)$ with alphabet $\mathcal{X}$ can be represented by a Bayes network $G' = (V', D')$ with $V' = V$ and $\mathcal{X}' = \mathcal{X}$.

- Choose a total ordering of $V' = V$ and write μ in terms of conditional probabilities.

In general the resulting Bayes network is dense.

Bayes networks are as powerful as factor graphs

A Bayes network $G = (V, D)$ with alphabet $\mathcal{X}$ can be represented by a factor graph model on $G' = (V', F', E')$ with $V' = V$, $|F'| = |V|$, $|E'| = |D| + |V|$, $\mathcal{X}' = \mathcal{X}$.

- Represent by a factor node each CPT.

A factor model on $G = (V, F, E)$ with alphabet $\mathcal{X}$ can be represented by a Bayes network $G' = (V', D')$ with $V' = V$ and $\mathcal{X}' = \mathcal{X}$.

- Choose a total ordering of $V' = V$ and write μ in terms of conditional probabilities.

In general the resulting Bayes network is dense.

Bayes networks are as powerful as factor graphs

A Bayes network $G = (V, D)$ with alphabet $\mathcal{X}$ can be represented by a factor graph model on $G' = (V', F', E')$ with $V' = V$, $|F'| = |V|$, $|E'| = |D| + |V|$, $\mathcal{X}' = \mathcal{X}$.

- Represent by a factor node each CPT.

A factor model on $G = (V, F, E)$ with alphabet $\mathcal{X}$ can be represented by a Bayes network $G' = (V', D')$ with $V' = V$ and $\mathcal{X}' = \mathcal{X}$.

- Choose a total ordering of $V' = V$ and write μ in terms of conditional probabilities.

In general the resulting Bayes network is dense.

Bayes networks are as powerful as factor graphs

A Bayes network $G = (V, D)$ with alphabet $\mathcal{X}$ can be represented by a factor graph model on $G' = (V', F', E')$ with $V' = V$, $|F'| = |V|$, $|E'| = |D| + |V|$, $\mathcal{X}' = \mathcal{X}$.

- Represent by a factor node each CPT.

A factor model on $G = (V, F, E)$ with alphabet $\mathcal{X}$ can be represented by a Bayes network $G' = (V', D')$ with $V' = V$ and $\mathcal{X}' = \mathcal{X}$.

- Choose a total ordering of $V' = V$ and write μ in terms of conditional probabilities.

In general the resulting Bayes network is dense.

Bayes networks with observed variables

$$\begin{aligned} V &= H \cup O\,, \\ x &= (x_i)_{i\in H} = (x_1,\dots,x_n) = (\text{ Hidden Variables })\,, \\ y &= (y_i)_{i\in O} = (y_1,\dots,y_m) = (\text{ Observed Variables }) \end{aligned}$$

$$\mu(x,y) = \prod_{i\in H} \mu(x_i | x_{\pi(i)\cap H}, y_{\pi(i)\cap O}) \prod_{i\in O} \mu(y_i | x_{\pi(i)\cap H}, y_{\pi(i)\cap O})$$

Of interest $\mu_y(x) = \mu(x|y)$

Bayes networks with observed variables

$$\begin{aligned} V &= H \cup O\,, \\ x &= (x_i)_{i\in H} = (x_1,\dots,x_n) = (\text{ Hidden Variables })\,, \\ y &= (y_i)_{i\in O} = (y_1,\dots,y_m) = (\text{ Observed Variables }) \end{aligned}$$

$$\mu(x,y) = \prod_{i\in H} \mu(x_i|x_{\pi(i)\cap H}, y_{\pi(i)\cap O}) \prod_{i\in O} \mu(y_i|x_{\pi(i)\cap H}, y_{\pi(i)\cap O})$$

Of interest $\mu_y(x) = \mu(x|y)$

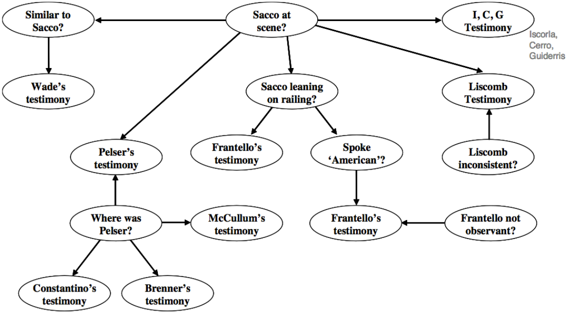
Bayes networks with observed variables

$$
\begin{aligned}
V &= H \cup O\,, \\
x &= (x_i)_{i\in H} = (x_1,\dots,x_n) = (\text{ Hidden Variables })\,, \\
y &= (y_i)_{i\in O} = (y_1,\dots,y_m) = (\text{ Observed Variables })
\end{aligned}
$$

$$
\mu(x,y) = \prod_{i\in H} \mu(x_i|x_{\pi(i)\cap H}, y_{\pi(i)\cap O}) \prod_{i\in O} \mu(y_i|x_{\pi(i)\cap H}, y_{\pi(i)\cap O})
$$

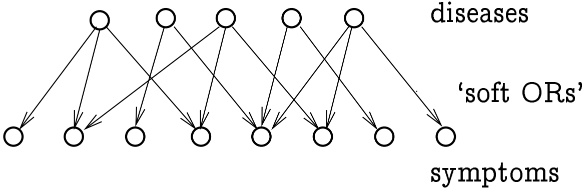
Of interest $\mu_y(x) = \mu(x|y)$

Example 1: Forensic science

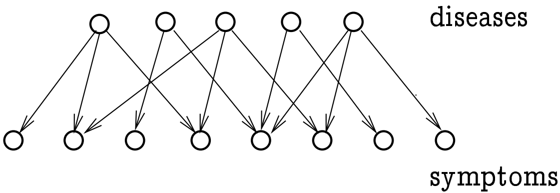


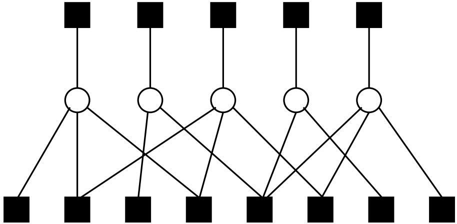
[Kadane, Shum, *A probabilistic analysis of the Sacco and Vanzetti evidence*, 1996]
[Taroni et al., *Bayesian Networks and Probabilistic Inference in Forensic Science*, 2006]

Example 2: Diagnostic network

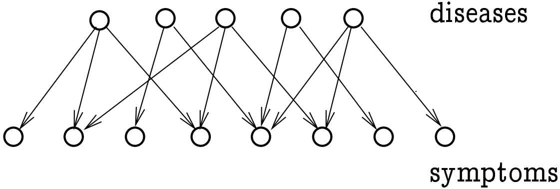


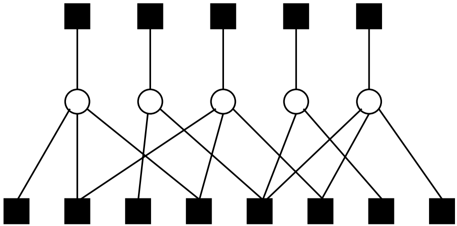
Example 2: Diagnostic network





Bayes networks are as powerful as factor graphs

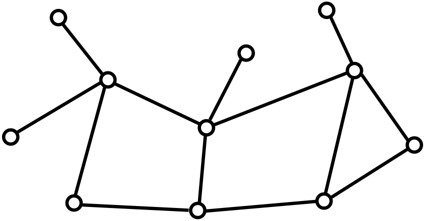




With observed variables graph remains unchanged

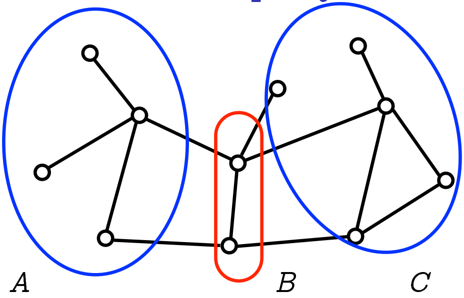
Markov property

Global Markov Property



$$\mu(x) = \mu(x_1, x_2, \ldots, x_n)$$

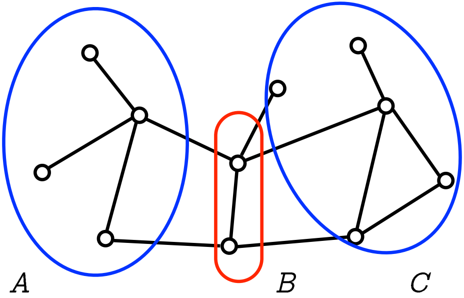
Global Markov Property



Definition

Let $A \cup B \cup C$ be a partition of V. We say that B separates A from C if any path starting in A and terminating in C has at least one node in B.

Global Markov Property



Definition

The probability distribution μ over $\mathcal{X}^V$ satisfies the global Markov property on G if for any partition $V = A \cup B \cup C$ such that B separates A from C,

$$\mu(x_A, x_C | x_B) = \mu(x_A | x_B)\, \mu(x_C | x_B)$$

The most general

Theorem (Hammersley, Clifford, 1971)

Let $\mu(\,\cdot\,)$ be a probability distribution on $\mathcal{X}^V$, and $G = (V, E)$ be a graph such that

- *μ is Markov with respect to G.*
- *$\mu(x) > 0$ for all $x \in \mathcal{X}^V$.*
- *G does not contain triangles.*

Then μ is a pairwise graphical model on G.

Proof sketch (Grimmett, Bull. London Math. Soc. 1973)

For every $S \subseteq V$, define

$$\widetilde{\psi}_S(x_S) \equiv \prod_{U \subseteq S} \mu(x_U, 0_{V \setminus U})^{(-1)^{|S \setminus U|}}$$

Example: For $S = \{i, j\}$ (not necessarily edge) let $\mu_{ij,+}(\,\cdot\,) \equiv \mu(\,\cdot\,, 0_{V \setminus \{i,j\}})$

$$\widetilde{\psi}_{ij}(x_i, x_j) = \mu_{ij,+}(x_i, x_j)\, \mu_{ij,+}(x_i, 0)^{-1} \mu_{ij,+}(0, x_j)^{-1}\, \mu_{ij,+}(0, 0)\,.$$

Exercise: If $V = \{i, j\} \ldots$

Proof sketch (Grimmett, Bull. London Math. Soc. 1973)

For every $S \subseteq V$, define

$$\widetilde{\psi}_S(x_S) \equiv \prod_{U \subseteq S} \mu(x_U, 0_{V \setminus U})^{(-1)^{|S \setminus U|}}$$

Example: For $S = \{i, j\}$ (not necessarily edge) let
$\mu_{ij,+}(\,\cdot\,) \equiv \mu(\,\cdot\,, 0_{V \setminus \{i,j\}})$

$$\widetilde{\psi}_{ij}(x_i, x_j) = \mu_{ij,+}(x_i, x_j)\, \mu_{ij,+}(x_i, 0)^{-1} \mu_{ij,+}(0, x_j)^{-1}\, \mu_{ij,+}(0, 0)\,.$$

Exercise: If $V = \{i, j\} \ldots$

Proof sketch (Grimmett, Bull. London Math. Soc. 1973)

For every $S \subseteq V$, define

$$\widetilde{\psi}_S(x_S) \equiv \prod_{U \subseteq S} \mu(x_U, 0_{V\setminus U})^{(-1)^{|S\setminus U|}}$$

Example: For $S = \{i, j\}$ (not necessarily edge) let
$\mu_{ij,+}(\,\cdot\,) \equiv \mu(\,\cdot\,, 0_{V\setminus\{i,j\}})$

$$\widetilde{\psi}_{ij}(x_i, x_j) = \mu_{ij,+}(x_i, x_j)\, \mu_{ij,+}(x_i, 0)^{-1} \mu_{ij,+}(0, x_j)^{-1}\, \mu_{ij,+}(0, 0)\,.$$

Exercise: If $V = \{i, j\}$...

Proof sketch (Grimmett, 1973)

- **Claim 1:** $\mu(x) = \mu(0_V) \prod_{S\subseteq V} \widetilde{\psi}_S(x_S)$

- **Claim 2:** For $S \neq \{i,j\} \in E$, $\{i\}$, $\widetilde{\psi}_S(x_S)$ =const.

Given μ, can construct G!

Proof sketch (Grimmett, 1973)

- **Claim 1:** $\mu(x) = \mu(0_V) \prod_{S \subseteq V} \widetilde{\psi}_S(x_S)$
- **Claim 2:** For $S \neq \{i,j\} \in E$, $\{i\}$, $\widetilde{\psi}_S(x_S) =$const.

Given μ, can construct G!

Useful facts

Fact

$$\sum_{K\subseteq W} (-1)^{|K|} = \mathbb{I}(W = \emptyset)\,.$$

(Sum includes the empty set.)

Fact

If f is a set function, $W' \subseteq W$ proper subset, then

$$\sum_{K\subseteq W} (-1)^{|K|} f(K\cap W') = 0\,.$$

(Sum includes the empty set.)

If G contains triangles

$$\mu(x) = \frac{1}{Z} \prod_{\mathcal{C} \in \text{cliques}(G)} \psi_{\mathcal{C}}(x_{\mathcal{C}})\,.$$

Footnote #1

Definition (Global Markov)

For any partition $V = A \cup B \cup C$ such that B separates A from C,

$$\mu(x_A, x_C | x_B) = \mu(x_A | x_B) \, \mu(x_C | x_B)$$

Local Markov: Required only for $A = \{i\}$, $B = \partial i$, $C = V \setminus \{i\} \cup \partial i$.

Pairwise Markov: Required only for $A = \{i\}$, $B = V \setminus \{i, j\}$, $C = \{j\}$.

They are in fact equivalent

Obviously

$$(G) \Rightarrow (L) \Rightarrow (P)$$

Less obviously: (P) $\Rightarrow$ (G)

By induction over $s \equiv |V \setminus B|$:

- $s = 2$: Pairwise property.
- Assume (G) for any $|B|$ with $|V \setminus B| = s \geq 2$ and prove it for $|V \setminus B| = s + 1$.

WLOG take $A \cup B \cup C = V$ and $|A| \geq 2$.

Less obviously: (P) $\Rightarrow$ (G)

By induction over $s \equiv |V \setminus B|$:

- $s = 2$: Pairwise property.
- Assume (G) for any $|B|$ with $|V \setminus B| = s \geq 2$ and prove it for $|V \setminus B| = s + 1$.

WLOG take $A \cup B \cup C = V$ and $|A| \geq 2$.

Induction step: The Intersection Lemma

We write A–B–C if x_A is conditionally independent of x_C given x_B.

Lemma

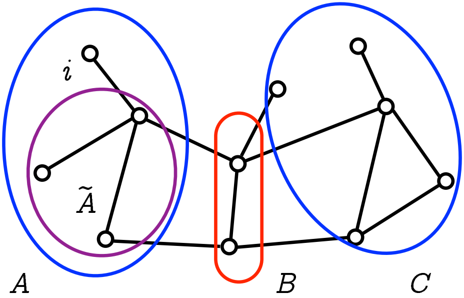
If μ is strictly positive and

$$A\text{–}(C \cup D)\text{–}B, \qquad A\text{–}(B \cup D)\text{–}C,$$

then

$$A\text{–}D\text{–}(B \cup C).$$

Induction step



$$A = \widetilde{A} \cup \{i\}$$

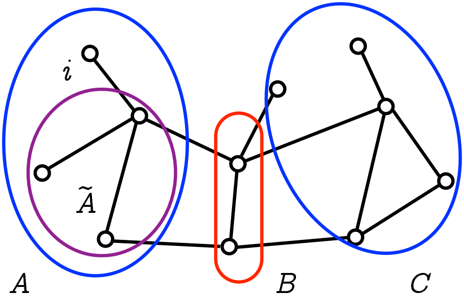
by induction assumption $\quad C \text{–} (B \cup i) \text{–} \widetilde{A}$

$$C \text{–} (B \cup \widetilde{A}) \text{–} \{i\}$$

By the intersection lemma

$$C \text{–} B \text{–} (\widetilde{A} \cup i) = A$$

Induction step



$$A = \tilde{A} \cup \{i\}$$

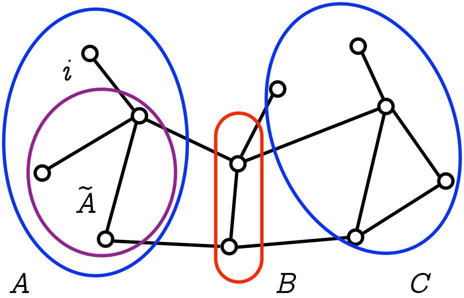
by induction assumption $\quad C – (B \cup i) – \tilde{A}$

$$C – (B \cup \tilde{A}) – \{i\}$$

By the intersection lemma

$$C – B – (\tilde{A} \cup i) = A$$

Induction step



$$A = \widetilde{A} \cup \{i\}$$

by induction assumption $\quad C–(B \cup i)–\widetilde{A}$

$$C–(B \cup \widetilde{A})–\{i\}$$

By the intersection lemma

$$C–B–(\widetilde{A} \cup i) = A$$

An interesting mathematical phenomenon

This is not true if G is infinite.

BTW: How do you define a MRF on an infinite graph?

An interesting mathematical phenomenon

This is not true if G is infinite.

BTW: How do you define a MRF on an infinite graph?

An interesting mathematical phenomenon

This is not true if G is infinite.

BTW: How do you define a MRF on an infinite graph?

Footnote #2

Where does the $(-1)^{|S \setminus U|}$ come from?

Möbius inversion formula

$$2^V = \{ \text{ subsets of } V \},$$

Theorem

Let $f, g : 2^V \to \mathbb{R}$. Then the following are equivalent

$$\begin{aligned} f(S) &= \sum_{U \subset S} g(U), \quad \textit{for all } S \subseteq V, \\ g(S) &= \sum_{U \subset S} (-1)^{|S \setminus U|} f(U), \quad \textit{for all } S \subseteq V. \end{aligned}$$

Proof: Exercise.

[see also G.C.Rota, Prob. Theor. Rel. Fields, 2 (1964) 340-368]

Relation with what we did

$$
\begin{aligned}
f(S) &= \log \mu(x_S, 0_{V\setminus S}), \\
g(S) &= \sum_{U\subset S} (-1)^{|S\setminus U|} \log \mu(x_U, 0_{V\setminus U}) = \tilde{psi}_S(x)
\end{aligned}
$$

Claim 1 in the proof ← Möbius