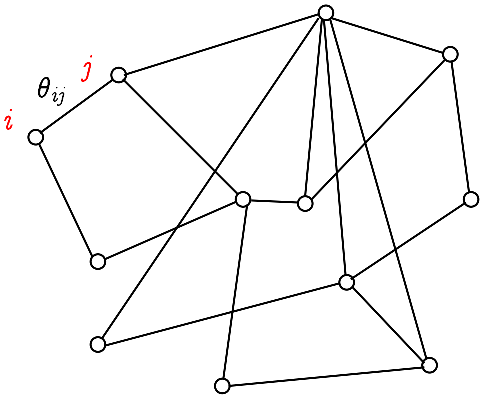
Inference and learning in Ising models

Andrea Montanari

Stanford University

January 4, 2012

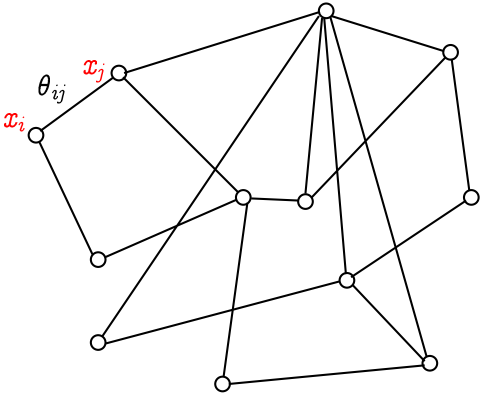
Definition



$$G = (V, E, \theta), \quad \theta : V \cup E \to \mathbb{R} \qquad \text{weighted graph}$$

$$\theta i \mapsto \theta(i) = \theta_i \,, \qquad \theta : (i,j) \mapsto \theta(i,j) = \theta_{i,j}$$

A probabilty distribution over $x \in \{+1, -1\}^V$.



$$\mu_{G,\theta}(x) = \frac{1}{Z_G(\theta)} \exp\Big\{ \sum_{(i,j)\in E} \theta_{ij} x_i x_j + \sum_{i \in V} \theta_i x_i \Big\}$$

Outline

- Why Ising models?
- Understanding $\mu_{G,\theta}$.
- Parameter and structure learning

Why Ising models?

Answer # 1: They pop up everywhere

Example: Coordination games on networks

Two-players coordination game

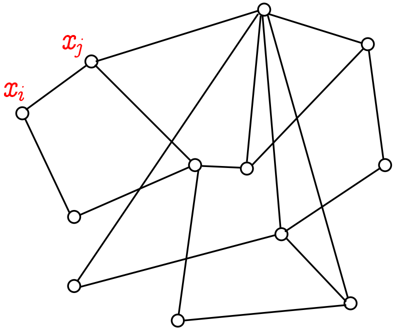
$$u_1(x_1, x_2) = u_2(x_2, x_1) = \begin{array}{c|c|c|} & + & - \\ \hline + & 5 & 4 \\ \hline - & 2 & 6 \\ \hline \end{array}$$

$5 > 2$, $6 > 4$

(2 Nash equilibria)

Coordination game on a network

A model for the evolution of social norms



$$U_i(x_i, x_{V\setminus i}) = \sum_{j\in\partial i} u_1(x_i, x_j)$$

($e^{\Theta(n)}$ Nash equilibria)

Idea: Study (noisy) best response dynamics

- M. Kandori, H. Mailath, F. Rob, *Learning, mutation, and long run equilibria in games*, Econometrica 1993
- H.P. Young, *The evolution of conventions*, Econometrica1993
- G. Ellison, *Learning, local interaction, and coordination*, Econometrica1993
- L. Blume, *The statistical mechanics of best-response strategy revision*, Games Econ. Behav. 1995
- H.P. Young, *The diffusion of innovation in social networks*, 2006
- A. Montanari, A. Saberi, *The spread of innovations in social networks*, PNAS 2010
- H.P. Young, *The dynamics of social innovation*, PNAS 2011

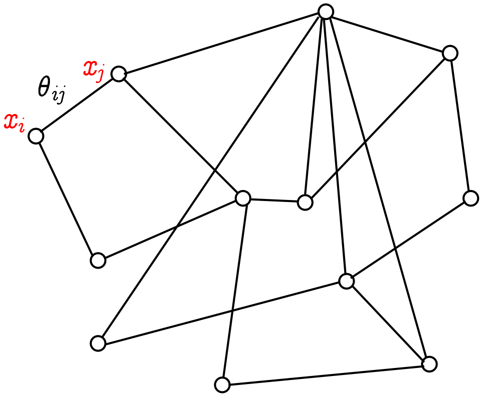
Noisy best response dynamics

Player i revises her strategy at Poisson times

$$p(x_i, x_{-i} \to x_i^{\text{new}}, x_{-i}) \propto \exp\Big\{\beta\, U_i(x_i^{\text{new}}, x_{-i})\Big\} \qquad x_i^{\text{new}} \in \{+1, -1\}$$

Logistic model (Blume, 1995)

Stationary distribution



$$\mu_{G,\theta}(x) = \frac{1}{Z_G(\theta)} \exp\Big\{ \sum_{(i,j)\in E} \theta x_i x_j + \sum_{i\in V} \theta_i x_i \Big\}, \qquad \theta_i = b\,|\partial i|\,.$$

$\theta,\ b > 0$

Answer # 2: Very rich family

Example: 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?

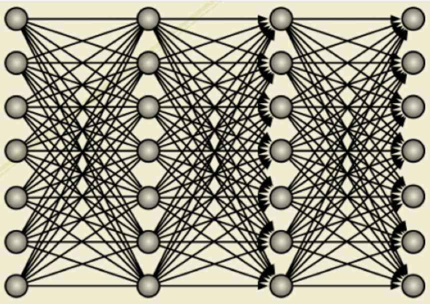
An attempt

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

What's the magic?

What's the magic?

Hidden variables



Image

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

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

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

Answer # 3: It is the most general. . .

- Pairwise binary graphical model.

- Binary and Markovian with respect to G.

Pairwise graphical model

$$G = (V, E), \qquad x = (x_i)_{i \in V}$$

$$\mu_G(x) = \frac{1}{Z_G(\psi)} \prod_{(i,j) \in E} \psi_{ij}(x_i, x_j)\,,$$

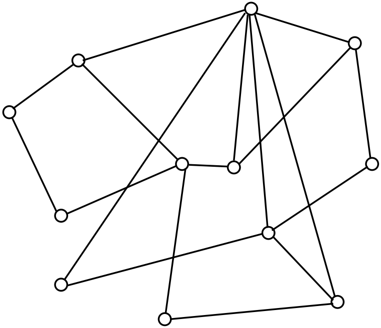
$Binary: \; x_i \in \{+1, -1\}$

Pairwise graphical model

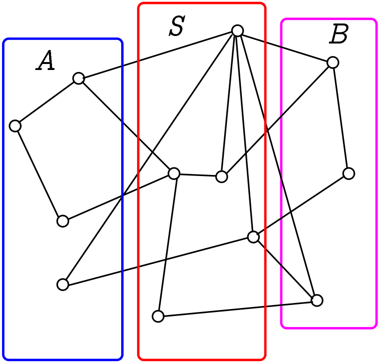
The most general function $\psi_{ij} : \{+1,-1\} \times \{+1,-1\} \to \mathbb{R}_+$

$$\log \psi_{i,j}(x_i, x_j) = c_0 + \tilde{\theta}_i x_i + \tilde{\theta}_j x_j + \theta_{ij} x_i x_j$$

Markov property on G



Markov property on G



Definition

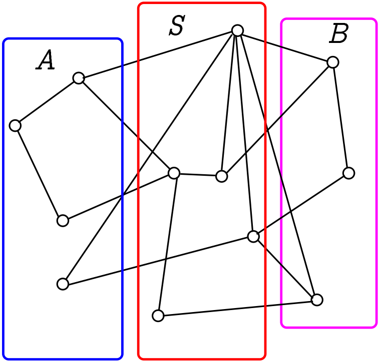
S *separates* A and B if every path on G departing from $a \in A$ and ending in $b \in B$ crosses S.

Markov property on G

Notation: For $A \subseteq V$

$$\mu(x_A) = \mu_A(x_A) = \sum_{x_{V\setminus A}} \mu(x_A, x_{V\setminus A}) = \mathbb{P}_\mu\{X_A = x_A\}$$

Markov property on G



Definition

$\mu(\,\cdot\,)$ over $\mathcal{X}^V$ is Markov with respect to G if, for any A, B, S such that S separates A from B, we have

$$\mu(x_A, x_B | x_S) = \mu(x_A | x_S) \mu(x_B | x_S)\,.$$

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.

Ising models are 'the only' binary distributions that are Markov wth respect to a graph G.

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

For every $S \subseteq V$, define

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

Example: For $S = \{i, j\}$ (not necessarily edge) let
$\mu_{ij,+}(\,\cdot\,) \equiv \mu(\,\cdot\,, (+1)_{V \setminus \{i,j\}})$

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

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

Proof sketch (Grimmett, 1973)

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

Given μ, can construct G!

Why to use exponentials?

$$\mu_{G,\theta}(x) = \frac{1}{Z_G(\theta)} \exp\Big\{ \sum_{(i,j)\in E} \theta_{ij} x_i x_j + \sum_{i\in V} \theta_i x_i \Big\}$$

- *Answer # 1:* My Ph.D. is in Physics.
- *Answer # 2:* I spend (50% of) my time in a Statistics department.

More seriously

Theorem

If $\mu(x) > 0$ for all $x \in \{+1, -1\}^V$, then the model parameters (G, θ) are uniquely determined by $M = (M^{(1)}, M^{(2)})$, $M^{(1)} = (M_i)_{i \in V}$, and $M^{(2)} = (M_{ij})_{i,j \in V}$ where

$$M_i \equiv \mathbb{E}_{\mu,\theta}\{x_i\}, \qquad M_{ij} \equiv \mathbb{E}_{\mu,\theta}\{x_i x_j\}.$$

Proof

Consider WLOG $G = K_n$, and define the free energy

$$\phi(\theta) \equiv \log Z(\theta) = \log \left\{ \sum_x e^{\sum_{(i,j} \theta_{i,j} x_i x_j + \sum_i \theta_i x_i} \right\}$$

Exercise: $\theta \mapsto \phi(\theta)$ is convex. Strictly convex if $\mu(x) > 0$ for all x
Hint: Compute the Hessian.

Proof

Consider

$$F(M,\theta) = \langle M,\theta\rangle - \phi(\theta)\,.$$

where $< M,\theta\rangle \equiv \sum_{i\in V}\theta_i M_i + \sum_{(ß,j)\in E}\theta_{ij} M_{ij}$.

- Stationarity conditions (in θ)

$$M_i = \mathbb{E}_\theta\{x_i\}\,,\qquad M_{ij} = \mathbb{E}_\theta\{x_i x_j\}$$

- Solution exists and is unique (by convexity).