ERGMs

A physicist's perspective

Pierre-André Noël

University of California, Davis

November 30th, 2015

ERGMs

- Exponential Random Graph Models.
  - Also called p^*.
- Special case of probable inference:
  - there is a **massive** number of possible graphs;
  - we know a **few** constraints;
  - and we seek the probability for each graph.
- Fill the gap with maximum entropy.
  - (Gap between small number of constraints and astronomical number of degrees of freedom.)
  - MaxEnt: no further assumptions than those explicitly stated in the constraints.
  - Result: exponential family, "E" in ERGMs.
  - Physicist: equilibrium statistical mechanics!

ERGMs

- Exponential Random Graph Models.
  - Also called p^*.
- Special case of probable inference:
  - there is a **massive** number of possible graphs;
  - we know a **few** constraints;
  - and we seek the probability for each graph.
- Fill the gap with maximum entropy.
  - (Gap between small number of constraints and astronomical number of degrees of freedom.)
  - MaxEnt: no further assumptions than those explicitly stated in the constraints.
  - Result: exponential family, "E" in ERGMs.
  - Physicist: equilibrium statistical mechanics!

ERGMs

- Exponential Random Graph Models.
  - Also called p^*.
- Special case of probable inference:
  - there is a **massive** number of possible graphs;
  - we know a **few** constraints;
  - and we seek the probability for each graph.
- Fill the gap with maximum entropy.
  - (Gap between small number of constraints and astronomical number of degrees of freedom.)
  - MaxEnt: no further assumptions than those explicitly stated in the constraints.
  - Result: exponential family, "E" in ERGMs.
  - Physicist: equilibrium statistical mechanics!

ERGMs

- Exponential Random Graph Models.
  - Also called p^*.
- Special case of probable inference:
  - there is a **massive** number of possible graphs;
  - we know a **few** constraints;
  - and we seek the probability for each graph.
- Fill the gap with maximum entropy.
  - (Gap between small number of constraints and astronomical number of degrees of freedom.)
  - MaxEnt: no further assumptions than those explicitly stated in the constraints.
  - Result: exponential family, "E" in ERGMs.
  - Physicist: equilibrium statistical mechanics!

ERGMs

- Exponential Random Graph Models.
  - Also called p^*.
- Special case of probable inference:
  - there is a **massive** number of possible graphs;
  - we know a **few** constraints;
  - and we seek the probability for each graph.
- Fill the gap with maximum entropy.
  - (Gap between small number of constraints and astronomical number of degrees of freedom.)
  - MaxEnt: no further assumptions than those explicitly stated in the constraints.
  - Result: exponential family, "E" in ERGMs.
  - Physicist: equilibrium statistical mechanics!

ERGMs

- Exponential Random Graph Models.
  - Also called p^*.
- Special case of probable inference:
  - there is a **massive** number of possible graphs;
  - we know a **few** constraints;
  - and we seek the probability for each graph.
- Fill the gap with maximum entropy.
  - (Gap between small number of constraints and astronomical number of degrees of freedom.)
  - MaxEnt: no further assumptions than those explicitly stated in the constraints.
  - Result: exponential family, "E" in ERGMs.
  - Physicist: equilibrium statistical mechanics!

ERGMs

- Exponential Random Graph Models.
  - Also called p^*.
- Special case of probable inference:
  - there is a **massive** number of possible graphs;
  - we know a **few** constraints;
  - and we seek the probability for each graph.
- Fill the gap with maximum entropy.
  - (Gap between small number of constraints and astronomical number of degrees of freedom.)
  - MaxEnt: no further assumptions than those explicitly stated in the constraints.
  - Result: exponential family, "E" in ERGMs.
  - Physicist: equilibrium statistical mechanics!

Contents

Source materials

- E.T. Jaynes. *Probability Theory: The Logic of Science.* Editor: G.L. Bretthorst. Cambridge University Press (2003).
  - For inference in general.
  - Here mostly chapter 11.
  - Hereafter known as [Jaynes].
- J. Park and M.E.J. Newman. *Statistical mechanics of networks.* Physical Review E **70**, 066117 (2004).
  - For the ERGMs part.
  - Hereafter known as [Park&Newman].

Outline

The 3-levels problem

- There are 3 types of gizmos: A, B, and C.
- Gizmo costs: $\epsilon_A = 1$, $\epsilon_B = 2$, $\epsilon_C = 3$.
- $N = 4$ gizmos purchased, total cost $U = 6$.
- $x_j \in \{A, B, C\}$ is the type of the j-th gizmo.
- State: $\mathbf{x} = (x_1, x_2, x_3, x_4)$.
- What is $\mathbb{P}(\mathbf{x})$?
- e.g., $\mathbb{P}\underbrace{(A, A, B, C)}_{1+1+2+3=7} = 0$, $\mathbb{P}\underbrace{(A, A, A, C)}_{1+1+1+3=6} \neq 0$

The 3-levels problem

- There are 3 types of gizmos: A, B, and C.
- Gizmo costs: $\epsilon_A = 1$, $\epsilon_B = 2$, $\epsilon_C = 3$.
- $N = 4$ gizmos purchased, total cost $U = 6$.
- $x_j \in \{A, B, C\}$ is the type of the j-th gizmo.
- State: $\mathbf{x} = (x_1, x_2, x_3, x_4)$.
- What is $\mathbb{P}(\mathbf{x})$?
- e.g., $\mathbb{P}\underbrace{(A, A, B, C)}_{1+1+2+3=7} = 0$, $\mathbb{P}\underbrace{(A, A, A, C)}_{1+1+1+3=6} \neq 0$

- $\mathbb{P}(\mathbf{x}) = 0$ unless $\sum_j \epsilon_{x_j} = 6$.
- Other cases $\mathbf{x} \in \{A, B, C\}^4$ are equiprobable.
- List the possibilities and count!

x_1	x_2	x_3	x_4	$\sum_j \epsilon_{x_j}$
~~A~~	~~A~~	~~A~~	~~A~~	~~4~~
~~A~~	~~A~~	~~A~~	~~B~~	~~5~~
A	A	A	C	6
~~A~~	~~A~~	~~B~~	~~A~~	~~5~~
A	A	B	B	6
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
~~C~~	~~C~~	~~C~~	~~C~~	~~12~~

- Now say $N = 4\,000\,000$ and $U = 6\,000\,000$.
  - Instead of $N = 4$ and $U = 6$.
- Here can still do combinatorics, but...
- This approach cares about irrelevant details.
  - e.g., number of B is even if N and U are even.
  - In general, partition problem: #P-complete.
- Much simpler problem with soft constraints.
  - Soft constraints only satisfied **on average**.
  - i.e., $\mathbb{E}\left(\sum_j \epsilon_{x_j}\right) = \sum_{\mathbf{x}} \mathbb{P}(\mathbf{x}) \sum_j \epsilon_{x_j} = U$ **(soft)**
  - instead of $\mathbb{P}(\mathbf{x}) = 0$ if $\sum_j \epsilon_{x_j} \neq U$ **(hard)**
- (*In physics, this is justified by a "heat bath", but here there is no need for such things.*)

Soft 3-levels problem, $N = 1$

- Hard constraint: $x \in \{A, B, C\}$
  - Therefore: $\mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) = 1$
- Soft constraint: $\mathbb{P}(A) + 2\,\mathbb{P}(B) + 3\,\mathbb{P}(C) = U$
  - With real-valued U such that $1 \leq U \leq 3$.
- Solution:

$$\mathbb{P}(A) = p_A$$
$$\mathbb{P}(B) = 3 - U - 2p_A$$
$$\mathbb{P}(C) = U - 2 + p_A$$

- There is still one unknown left!

- Suppose $U = 2$.
  - Then p_A can be anywhere in interval $\left[0, \frac{1}{2}\right]$!
  - But intuition tells us that $p_A = \frac{1}{3}$ is the "best", because it gives $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{3}$.
  - Without any reasons to prefer A, B or C, we make the distribution as "flat" as possible.
- Now suppose $U > 2$.
  - If $U \leq \frac{5}{3}$, the solution $p_A = \frac{1}{3}$ is still possible.
  - However, that would give $\mathbb{P}(B) < \frac{1}{3} > \mathbb{P}(C)$.
  - Nothing special about $\mathbb{P}(A)$, no reason why this one should stay $\frac{1}{3}$ and not the others.
  - Intuition: $p_A < \frac{1}{3}$, but by how much?
- Make the distribution as "flat" as possible?
  - Universal measure of flatness? Entropy!

Outline

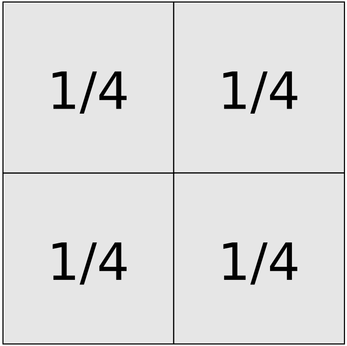
The 3-levels problem

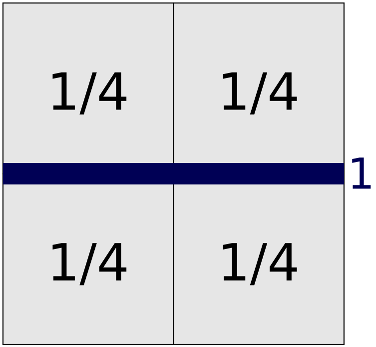
Entropy 101

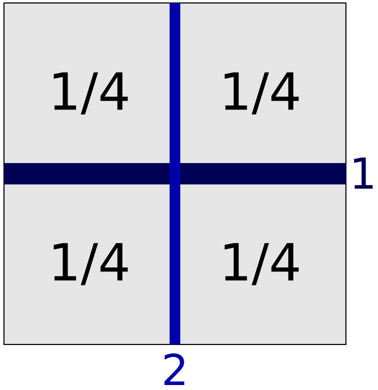
Statistical mechanics of anything

ERGMs at last!

A physicist's view of MCMC

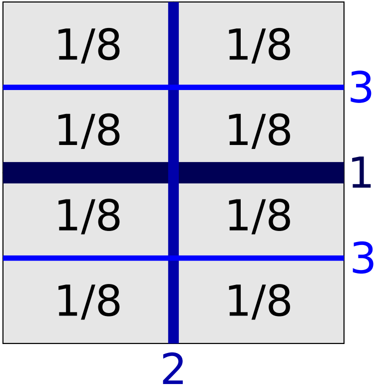

1/4
1/4
1/4
1/4


1/4
1/4
1
1/4
1/4

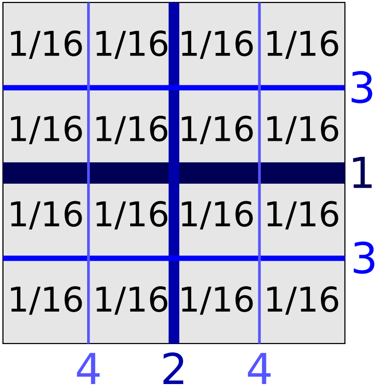


1/4
1/4
1/4
1/4
1
2

1/8	1/8
1/8	1/8
1/8	1/8
1/8	1/8

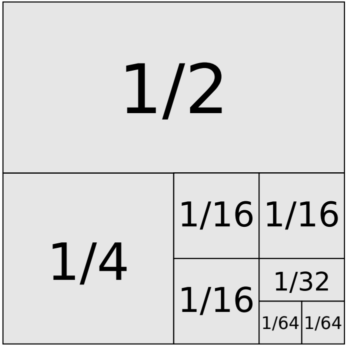

1/8
1/8
3
1/8
1/8
1
1/8
1/8
3
1/8
1/8
2

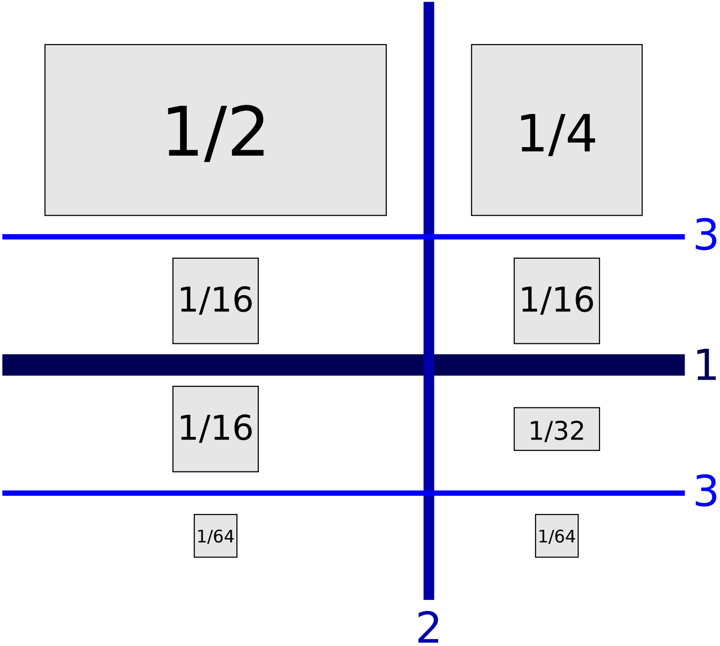
1/16	1/16	1/16	1/16
1/16	1/16	1/16	1/16
1/16	1/16	1/16	1/16
1/16	1/16	1/16	1/16


1/16 1/16 1/16 1/16
1/16 1/16 1/16 1/16
1/16 1/16 1/16 1/16
1/16 1/16 1/16 1/16
3
1
3
4
2
4

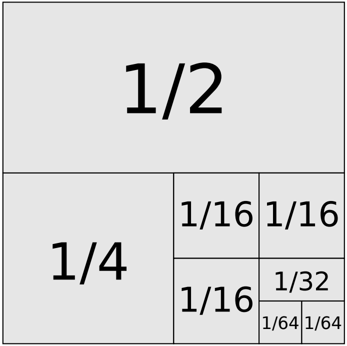
Equiprobable choices

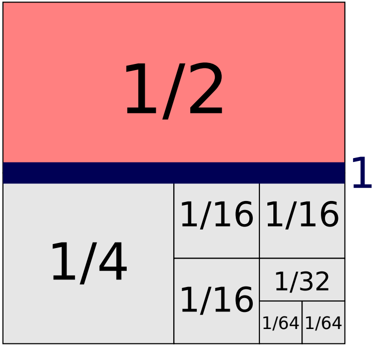
- If there are $N = 2^q$ equiprobable choices,
  - then you need $q = \log_2 N$ yes/no questions.
- Each value has probability $p = \frac{1}{N}$ to occur,
  - so we have $q = -\log_2 p$.


1/2
1/4
1/16
1/16
1/16
1/32
1/64
1/64

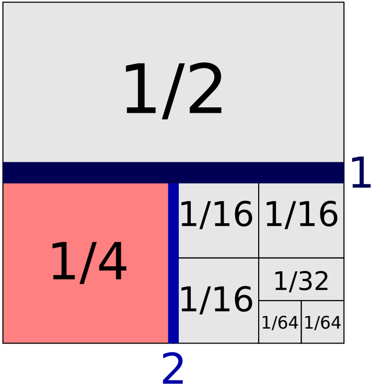


1/2
1/4
3
1/16
1/16
1
1/16
1/32
3
1/64
1/64
2

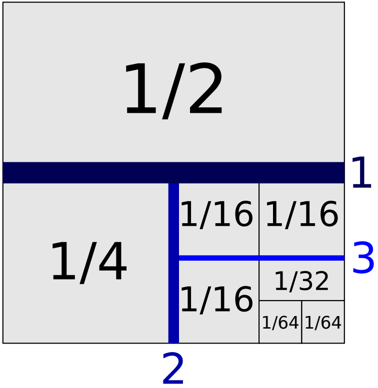

1/2
1/4
1/16
1/16
1/16
1/32
1/64
1/64



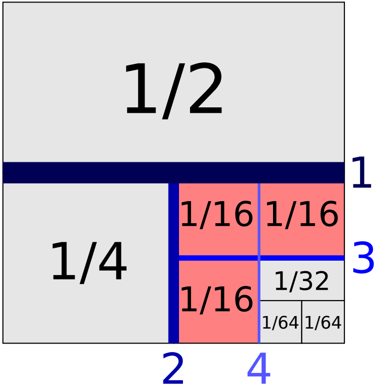
q	1
$\mathbb{P}(q)$	$\frac{1}{2}$
# discerned states	1
$\mathbb{P}(x)$	$\frac{1}{2}$



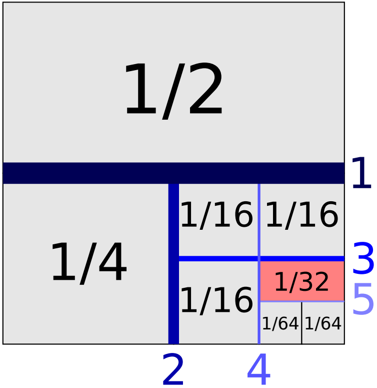
q	1	2
$\mathbb{P}(q)$	$\frac{1}{2}$	$\frac{1}{4}$
# discerned states	1	1
$\mathbb{P}(x)$	$\frac{1}{2}$	$\frac{1}{4}$



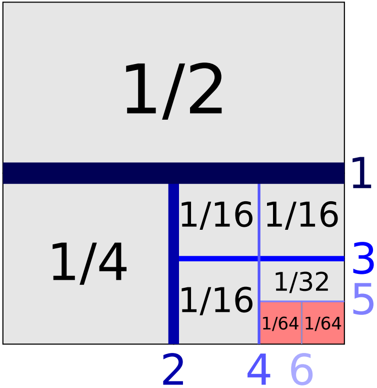
q	1	2	3
$\mathbb{P}(q)$	$\frac{1}{2}$	$\frac{1}{4}$	0
# discerned states	1	1	0
$\mathbb{P}(x)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$



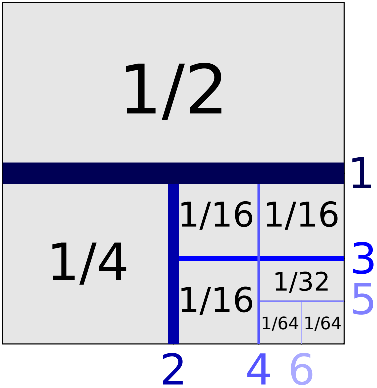
q	1	2	3	4
$\mathbb{P}(q)$	$\frac{1}{2}$	$\frac{1}{4}$	0	$\frac{3}{16}$
# discerned states	1	1	0	3
$\mathbb{P}(x)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$



q	1	2	3	4	5
$\mathbb{P}(q)$	$\frac{1}{2}$	$\frac{1}{4}$	0	$\frac{3}{16}$	$\frac{1}{32}$
# discerned states	1	1	0	3	1
$\mathbb{P}(x)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$

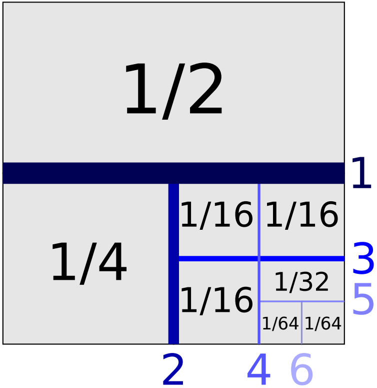


q	1	2	3	4	5	6
$\mathbb{P}(q)$	$\frac{1}{2}$	$\frac{1}{4}$	0	$\frac{3}{16}$	$\frac{1}{32}$	$\frac{2}{64}$
# discerned states	1	1	0	3	1	2
$\mathbb{P}(x)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{64}$



$$\mathbb{E}(q) = \sum_q \mathbb{P}(q) q \approx 2.6 < 3$$

q	1	2	3	4	5	6
$\mathbb{P}(q)$	$\frac{1}{2}$	$\frac{1}{4}$	0	$\frac{3}{16}$	$\frac{1}{32}$	$\frac{2}{64}$
# discerned states	1	1	0	3	1	2
$\mathbb{P}(x)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{64}$



$$\mathbb{E}(q) = \sum_q \mathbb{P}(q)q \approx 2.6 < 3$$

$$= -\sum_x \mathbb{P}(x) \log_2 \mathbb{P}(x)$$

$$\left(q(x) = -\log_2 \mathbb{P}(x)\right)$$

q	1	2	3	4	5	6
$\mathbb{P}(q)$	$\frac{1}{2}$	$\frac{1}{4}$	0	$\frac{3}{16}$	$\frac{1}{32}$	$\frac{2}{64}$
# discerned states	1	1	0	3	1	2
$\mathbb{P}(x)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{64}$

Information entropy $H(\mathbf{p})$

- $$H(\mathbf{p}) \equiv -\sum_x p_x \log_b p_x = -k \sum_x p_x \ln p_x$$
- Valid in general, not just for powers of $\frac{1}{2}$.
- If $b = 2$ (or $k = \frac{1}{\ln 2}$), then unit is *bits*.
  - Average number of yes/no questions required to learn the value of x (when optimum).
  - Easier to get an intuition of what it means.
- If $b = \mathrm{e}$ (or $k = 1$), then unit is *nats*.
  - Practical for analysis, because $\dfrac{\mathrm{d} \ln p}{\mathrm{d} p} = \dfrac{1}{p}$.
  - From now on, we use $b = \mathrm{e}$.

Properties and interpretation

$$H(\mathbf{p}) = -\sum_x p_x \ln p_x$$

- Measures how much we don't know about x.
- If $\mathbf{p}$ has N nonzero entries, then $H(\mathbf{p}) \leq \ln N$
  - $H(\mathbf{p}) = \ln N$ iff. all nonzero entries are $\frac{1}{N}$.
  - Maximal when $\mathbf{p}$ flat: a measure of "flatness".
- In fact, **the** measure of flatness!
  - The "right" $\mathbf{p}$ is the one that maximizes $H(\mathbf{p})$ while satisfying constraints (hard and soft).
  - Use all that you know, but only that.

Back to the 3-levels problem

$$
\begin{aligned}
\mathbb{P}(A) &= p_A \\
\mathbb{P}(B) &= 3 - U - 2p_A \\
\mathbb{P}(C) &= U - 2 + p_A
\end{aligned}
$$

$$
\begin{aligned}
H(p_A) = -p_A \ln p_A \\
&- (3 - U - 2p_A) \ln(3 - U - 2p_A) \\
&- (U - 2 + p_A) \ln(U - 2 + p_A)
\end{aligned}
$$

- Maximal $H(p_A)$ when $\frac{\mathrm{d}H(p_A)}{\mathrm{d}p_A} = 0$.
- Solve: $p_A = \frac{10 - 3U - \sqrt{4 - 4(U-2)^2}}{6}$.
    - To be sure, should also check that $\frac{\mathrm{d}^2 H(p_A)}{\mathrm{d}p_A^2} < 0$.
- A little underwhelming here...
    - but we can proceed the same way in general!

Outline

The 3-levels problem

Entropy 101

Statistical mechanics of anything

ERGMs at last!

A physicist's view of MCMC

Defining the general problem

- Given a set $\mathcal{X}$ of accessible states.
  - Strong constraint:
    $\mathbb{P}(x) = 0$ if $x \notin \mathcal{X}$, thus $\sum_{x \in \mathcal{X}} \mathbb{P}(x) = 1$.
- Given some $f_c(x)$ and ϕ_c for $1 \leq c \leq C$.
  - $f_c : \mathcal{X} \to \mathbb{R}$, a real function of x.
  - $\phi_c \in \mathbb{R}$, a real number.
  - Weak constraints:
    $$\sum_{x \in \mathcal{X}} \mathbb{P}(x) f_c(x) = \phi_c \quad \forall c \in \{1, 2, \cdots, C\}$$
- Find $\mathbf{p}$ such that $\mathbb{P}(x) = p_x \; \forall x \in \mathcal{X}$.
  - i.e., $\mathbf{p}$ maximizing $H(\mathbf{p}) = -\sum_{x \in \mathcal{X}} p_x \ln p_x$.

Lagrange multipliers

- Method that allows to extremize in a high-dimensional space under constraints.
  - Equivalent of $\frac{\mathrm{d}H(p_A)}{\mathrm{d}p_A} = 0$ earlier.

$$\frac{\partial}{\partial p_{x'}}\left[\overbrace{-\sum_{x\in\mathcal{X}} p_x \ln p_x}^{H(\mathbf{p})} + \theta\left(\overbrace{1-\sum_{x\in\mathcal{X}} p_x}^{\text{normalization}}\right) + \sum_{c=1}^{C}\lambda_c\left(\overbrace{\phi_c - \sum_{x\in\mathcal{X}} p_x f_c(x)}^{\mathbb{E}(f_c(x))=\phi_c}\right)\right] = 0 \quad \forall x' \in \mathcal{X}$$

- $\theta, \lambda_1, \cdots, \lambda_c$ are Lagrange multipliers.
  - θ enforces normalization (strong constraint).
  - $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_c)$ enforces weak constraints.
- θ and $\boldsymbol{\lambda}$ unknown for now, fix at the end.

- After differentiation, we get:

$$0 = -1 - \ln p_x - \theta - \sum_{c=1}^{C} \lambda_c f_c(x)$$

$$p_x = \frac{1}{\mathrm{e}^{\theta+1}} \exp\left(- \sum_{c=1}^{C} \lambda_c f_c(x) \right)$$

- Now fix θ by enforcing normalization.

$$\mathrm{e}^{\theta+1} = \sum_{x \in \mathcal{X}} \exp\left(- \sum_{c=1}^{C} \lambda_c f_c(x) \right) \equiv Z(\boldsymbol{\lambda})$$

- $Z(\boldsymbol{\lambda})$ is called the partition function.

- Now fix $\boldsymbol{\lambda}$ by enforcing weak constraints.

$$\begin{aligned}\phi_c &= \sum_{x\in\mathcal{X}} p_x f_c(x) \\ &= \sum_{x\in\mathcal{X}} \frac{1}{Z(\boldsymbol{\lambda})} \exp\left(-\sum_{c=1}^{C} \lambda_c f_c(x)\right) f_c(x) \\ &= -\frac{\partial \ln Z(\boldsymbol{\lambda})}{\partial \lambda_c} \quad \forall c \in \{1, \cdots, C\}\end{aligned}$$

- We note $\boldsymbol{\Lambda}(\boldsymbol{\phi})$ the value of $\boldsymbol{\lambda}$ solving this.
- An we are done!

In a nutshell

- Specify the problem through constraints.
  - Strong: $\mathbb{P}(x) = 0$ if $x \notin \mathcal{X}$
  - Weak: $\sum_{x \in \mathcal{X}} \mathbb{P}(x) f_c(x) = \phi_c \ \forall c \in \{1, 2, \cdots, C\}$
- Obtain the solution.
  - Define: $Z(\boldsymbol{\lambda}) \equiv \sum_{x \in \mathcal{X}} \exp\left(-\sum_{c=1}^{C} \lambda_c f_c(x)\right)$
  - Find $\boldsymbol{\Lambda}(\boldsymbol{\phi})$ such that: $F_c = -\left.\frac{\partial \ln Z(\boldsymbol{\lambda})}{\partial \lambda_c}\right|_{\boldsymbol{\lambda}=\boldsymbol{\Lambda}(\boldsymbol{\phi})}$
  - Solution:

$$\mathbb{P}(x) = \frac{1}{Z(\boldsymbol{\Lambda}(\boldsymbol{\phi}))} \exp\left(-\sum_{c=1}^{C} \Lambda_c(\boldsymbol{\phi}) f_c(x)\right)$$

Outline

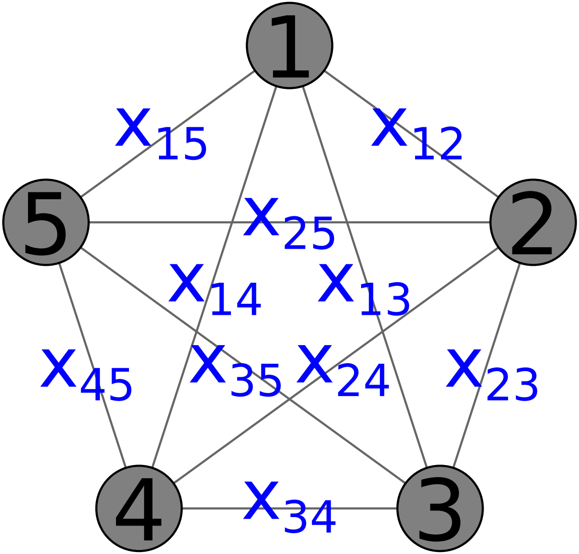
The 3-levels problem

Entropy 101

Statistical mechanics of anything

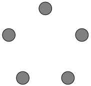
ERGMs at last!

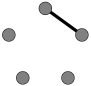
A physicist's view of MCMC

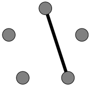


$$\mathbf{x} = (x_{12}, x_{13}, x_{14}, x_{15}, x_{23}, x_{24}, x_{25}, x_{34}, x_{35}, x_{45})$$

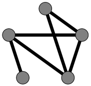
- Strong constraint: $\mathcal{X} = \{0,1\}^{10}$

$(0,0,0,0,0,0,0,0,0,0) \quad \Rightarrow$ 

$(1,0,0,0,0,0,0,0,0,0) \quad \Rightarrow$ 

$(0,1,0,0,0,0,0,0,0,0) \quad \Rightarrow$ 

$\vdots$

$(1,1,0,0,1,0,1,0,1,1) \quad \Rightarrow$ 

$\vdots$

- Weak constraints? The sky's the limit!
    - $f_1(\mathbf{x}) = \sum_{i<j} x_{ij}$ to get ϕ_1 edges (on average).

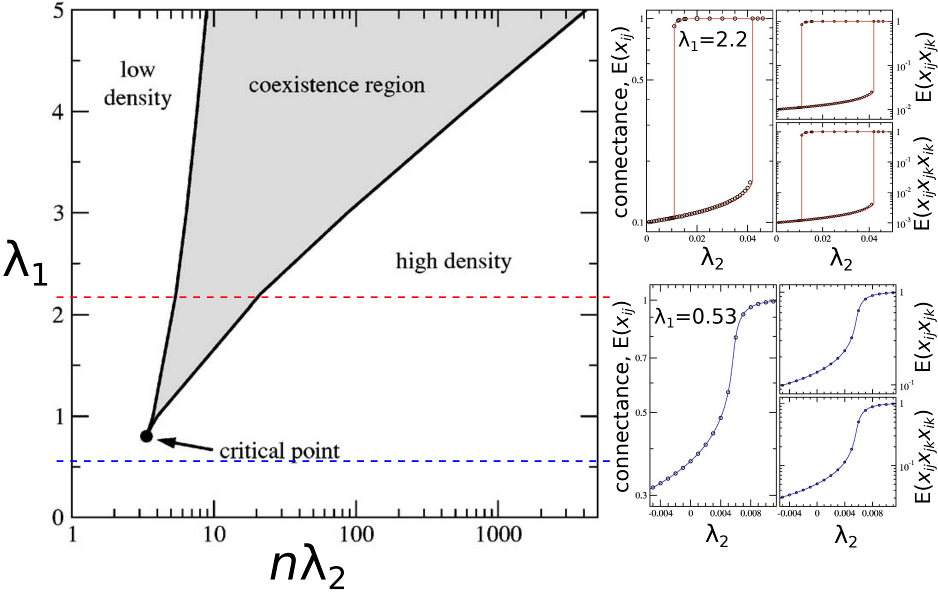
$$Z = \sum_{\mathbf{x}\in\mathcal{X}} e^{-\lambda_1 \sum_{i<j} x_{ij}} = (1 + e^{-\lambda_1})^{\binom{n}{2}}$$

$$\mathbb{P}(\mathbf{x}) = p^{f_1(\mathbf{x})}(1-p)^{\binom{n}{2} - f_1(\mathbf{x})}, \; p \equiv \frac{1}{e^{\Lambda_1(\phi_1)}+1} = \frac{\phi_1}{\binom{n}{2}}$$

    - Add $f_2(\mathbf{x}) = \sum_{i<j<k} x_{ij}x_{jk}x_{ik}$ to get ϕ_2 triangles.

$$Z = \sum_{\mathbf{x}\in\mathcal{X}} e^{-\lambda_1 \sum_{i<j} x_{ij}} e^{-\lambda_2 \sum_{i<j<k} x_{ij}x_{jk}x_{ik}} = woops \ldots$$

- When it starts to get interesting, no closed form for Z: we need to do approximations.



Park & M.E.J. Newman. PRE **72**, 026136 (2005).

Stat. mechanics of ERGMs

- It does exactly what you asked it to do.
  - Its not its fault that what you asked it to do is not what you want it to do.
    $\Rightarrow$ Clustering is more than just triangles.
- Lack of closed form for Z.
  - Approximations can be great (large graphs), once you figured out how to do it.
  - Like standard statistical mechanics, it can get pretty hard pretty quickly.
- But you can also do simulations!

Outline

Idea behind MCMC

- MCMC means Markov Chain Monte Carlo.
- Goal: sample a $\mathbf{x}$ from a distribution $p_{\mathbf{x}}$.
- Trick: Markov Chain $\mathbb{P}(\mathbf{x}^t|\mathbf{x}^{t-1})$ with $\pi_{\mathbf{x}} = p_{\mathbf{x}}$
  - For T large enough, $\mathbf{x}^T$ "sampled" from $p_{\mathbf{x}}$.
  - If wait more, another sample $\mathbf{x}^{2T}$, etc.
  - Necessary condition: $\sum_{\mathbf{x}'} \mathbb{P}(\mathbf{x}|\mathbf{x}')p_{\mathbf{x}'} = p_{\mathbf{x}}$
  - Sufficient: $\mathbb{P}(\mathbf{x}|\mathbf{x}')p_{\mathbf{x}'} = \mathbb{P}(\mathbf{x}'|\mathbf{x})p_{\mathbf{x}} \;\; \forall \mathbf{x}, \mathbf{x}' \in \mathcal{X}$
- Many different kind of MCMC.
  - Now see example of Metropolis–Hastings.

Metropolis–Hastings

$$p_{\mathbf{x}} = \frac{\exp\left(-\sum_c \lambda_c f_c(\mathbf{x})\right)}{Z(\boldsymbol{\lambda})}$$

$$\frac{p_{\mathbf{x}'}}{p_{\mathbf{x}}} = \exp\left(-\sum_c \lambda_c \left(f_c(\mathbf{x}') - f_c(\mathbf{x})\right)\right)$$

$$\frac{p_{\mathbf{x}'}}{p_{\mathbf{x}}} = \frac{\mathbb{P}(\mathbf{x}'|\mathbf{x})}{\mathbb{P}(\mathbf{x}|\mathbf{x}')} = \frac{Q(\mathbf{x}'|\mathbf{x})}{Q(\mathbf{x}|\mathbf{x}')} \frac{A(\mathbf{x}'|\mathbf{x})}{A(\mathbf{x}|\mathbf{x}')}$$

- "The" Metropolis–Hastings:
  - $Q(\mathbf{x}'|\mathbf{x}) = Q(\mathbf{x}|\mathbf{x}')$ and $A(\mathbf{x}'|\mathbf{x}) = \min\left(1, \frac{p_{\mathbf{x}'}}{p_{\mathbf{x}}}\right)$
  - Can check that solves the above.

- Bad things: proposal function $Q(\mathbf{x}|\mathbf{x}')$
  - Can have *major* impact on convergence time.
  - One more thing to tweak; a form of art.
- Good things: observable functions $\mathbf{f}(\mathbf{x})$
  - If $\mathbf{f}(\mathbf{x})$ local and if $\mathbf{x}$ very similar to $\mathbf{x}'$
  - then most terms in $\mathbf{f}(\mathbf{x}') - \mathbf{f}(\mathbf{x})$ cancel out.
  - Only need terms involving changed entries x_{ij}.
- Okay things: multipliers $\boldsymbol{\lambda}$ and constraints $\boldsymbol{\phi}$
  - If you have some guess for $\boldsymbol{\lambda}$, start with that.
  - Otherwise, just start with $\boldsymbol{\lambda} = \mathbf{0}$.
  - Now let algorithm run, but not for too long.
  - Check if $\mathbf{f}(\mathbf{x})$ has gotten closer to $\boldsymbol{\phi}$.
  - If not, tune $\boldsymbol{\lambda}$ to drive $\mathbf{x}$ that way. Repeat.
  - Final fine tuning: use $\mathbb{E}\big(\mathbf{f}(\mathbf{x})\big)$ instead of $\mathbf{f}(\mathbf{x})$.

Conclusion

- General inference problem:
  - Acknowledge the magnitude of how much we don't know (by maximizing entropy),
  - while accounting for all the constraints that we know (enforced by Lagrange multipliers).
- When done for graphs, gives ERGMs.

- Analysis (equilibrium stat. mech.):
    - Good for big picture, phase transitions, ...
    - When it gets messy, it gets messy.
- Simulations (MCMC):
    - Can do anything; don't care about messiness.
    - Choice of $Q(\mathbf{x}|\mathbf{x}')$ is an extra hurdle.
- Both:
    - Does exactly what you ask it to do.
    - It can't guess what you mean.

"There is on average ϕ_2 triangles."

is different from

"There is approximately ϕ_2 triangles."

Thank you!

(And now please welcome someone who actually knows what he is talking about!)