Recapping Inference Algs

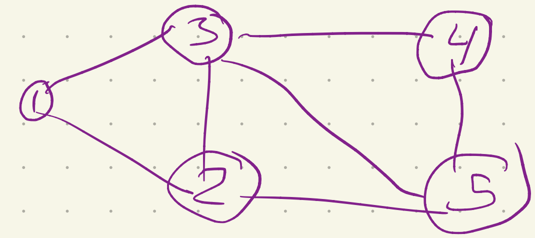
Marginalization	Maximization	Exact/Apx	Which GMs?	C.C.
Elimination Alg		Exact	Any G	$O(\lvert\chi\rvert^{TW})$ or worse
Sum-product = Belief Prop	Max-product	Apx	Pairwise MRFs	$O(\lvert\chi\rvert^{2})$
Sum-product on FGs = Belief Prop	Max-product on FGs	Apx	any FG	$O(\lvert\chi\rvert^{\text{max-deg}})$
Elimination Alg on junction trees		Exact	Junction trees	$O(\lvert\chi\rvert^{TW})$ or worse

Junction Tree Alg : Elim on A junction $\boxed{\text{tree}}$ → exact

- Fixing the tree, inference is easy using most methods
- Similar to elimination alg [exact, ordering]
- the data structure is meant to support efficient elimination

- MRF ⟶ Clique tree
  [non-unique]

$\boxed{M}$



One example clique tree for M

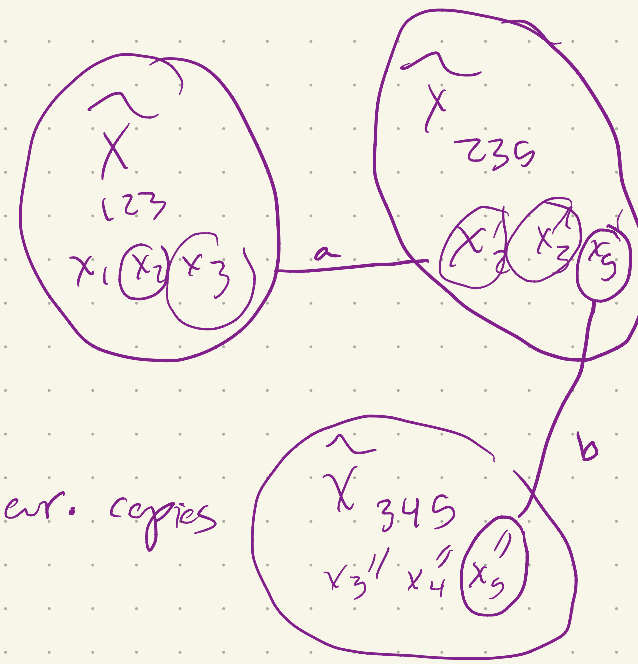
$$P(X) \propto f_{123}\, f_{235}\, f_{345}$$

– Create a joint node for each clique
  $\tilde{x}_C \in \mathcal{X}^{|C|}$

  - each node has a local copy of its vars

$$\tilde{x}_{123} = (x_1, x_2, x_3)$$
$$\tilde{x}_{235} = (x_2', x_3', x_5')$$
$$\tilde{x}_{345} = (x_3'', x_4'', x_5'')$$

  - Assign edges to form a tree
    & ensure consistency across var. copies



$$\tilde{P}(\tilde{x}_{123}, \tilde{x}_{235}, \tilde{x}_{345}) = \frac{1}{Z} f_{123}(\tilde{x}_{123}) \cdot f_{235}(\tilde{x}_{235}) f_{345}(\tilde{x}_{345}) \cdot$$

$$\mathbb{I}([\tilde{x}_{123}]_2 = [\tilde{x}_{235}]_1) \cdot \mathbb{I}([\tilde{x}_{123}]_3 = [\tilde{x}_{235}]_2) \cdot \mathbb{I}([\tilde{x}_{235}]_3 = [\tilde{x}_{345}]_3)$$

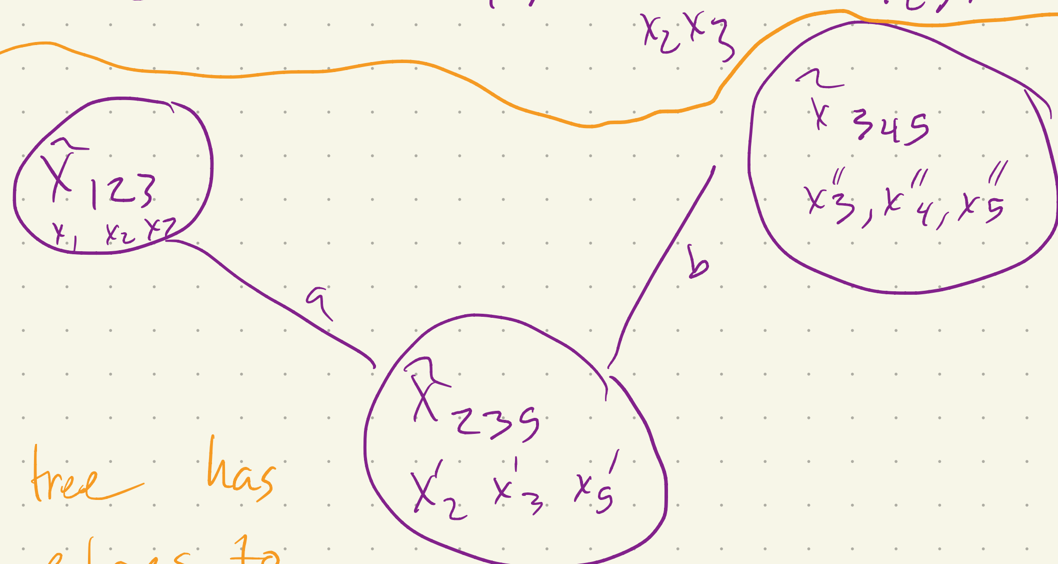
$$\mathbb{I}([\tilde{x}_{345}]_1 = [x_{235}]_2)$$

Consistency checks across edges

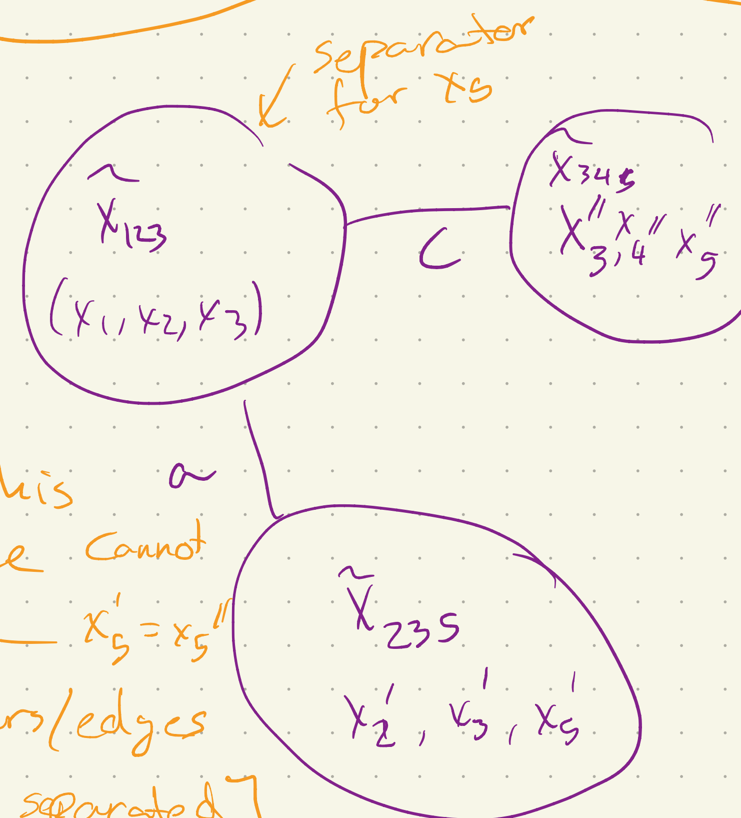
$$:= f_a(\tilde{x}_{123}, \tilde{x}_{235}) \cdot f_b(\tilde{x}_{345})$$

* If all local vars can be made consistent, we have global consistency

$$P(x_1) = \sum_{x_2 x_3} \tilde{P}(\tilde{x}_{123})$$



This tree has suff edges to guarantee global consistency



This tree cannot ensure $x_5' = x_5''$

[the factors/edges a+c are separated]

Q: When will a tree be globally consistent?

Def **Junction tree property**

For a MRF $G=(V,E)$ with $\mathcal{C}$ the set of maximal cliques

a tree T over $\mathcal{C}$, node $i \in V$ satisfies JTP if

all cliques containing i form a connected subtree $\in T$.

Def T is a junction tree if $\forall i \in V$ it has JTP.

Q: When will G have a JT? [existence]

A: when G is chordal (any ≥ 4 loop has a chord)

Q How to find a JT for G? ① [construction]

Q If G isn't chordal, is all hope lost? ②

① Algo sketch to find JT given $G=(V,E)$, $\mathcal{C}$ the set of max cliques

- Consider K_n on $\mathcal{C}$
- Assign weights on edges base on # shared vars $\Leftrightarrow$ 2 cliques
- Find max-wt spanning tree [using, eg, Kruskal's alg]

Claim A clique tree T is a junction tree $\Longleftrightarrow$ it's a max wt spanning tree

$$w(T) = \sum_{(C_e, C_K) \in T} |C_e \cap C_K|$$

$\forall$ trees T

$$= \sum_{(C_e, C_K) \in T} \sum_{i \in V} \mathbb{I}[i \in C_e \cap i \in C_K]$$

$$= \sum_{i \in V} \sum_{(C_\ell, C_k) \in T} \mathbb{I}[i \in C_\ell \cap i \in C_k]$$

$$\leq \sum_{i \in V} \sum_{(C_\ell, C_k) \in T} m_i - 1$$

m_i: # of cliques containing node i

Since cliques containing i is a subtree / forest on m_i nodes

$$\leq \sum_{i \in V} (m_i - 1)$$

$=$ only when subgraph on i's cliques is connected

$\Rightarrow$ T is a junction tree $\Longleftrightarrow$ T achieves max wt

D

② If G isn't chordal, what can you do?

Algo Sketch

- Choose an elim ordering
- Find reconstituted graph [chordal] $\Rightarrow$ Junction tree
- Construct complete clique graph
  Find max-wt spanning tree

We can run any exact algo on this JT and get exact inference.

Diff elim orderings $\to$ different [max clique sizes] / (complexities)

@ best $O(|\mathcal{X}|^{\text{tree width}})$

How good is BP for $|\mathcal{X}| < \infty$?

- When G is a tree, it's exact
- $\exists$ exactly $\underline{1}$ loop, it converges but might be wrong [Weiss 2000]
- Max-wt matching prob $\Rightarrow$ converges + exact [Bayati, Shah, Sharma 2005]
- In the limit of a large graph density estimation provides an asymptotic perf est.

Useful for

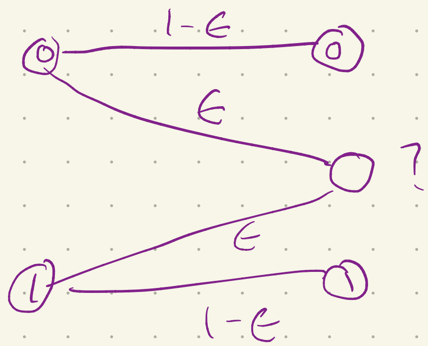
- Analysis of Compressed sensing
  $$\min_{x \in \mathbb{R}^d} \|Ax - b\|^2 + \lambda \|x\|_1$$
  ↳ Underdetermined system → sparse
- Analysis of LDPC decoders, first provable capacity-achieving codes [Luby '98]
- Analysis of near-opt crowdsourcing [Karger oh Shah 2011]

Work through density estimation for binary erasure channel (BEC)
[gives an example of analyzing perf of BP on large, sparse RGs]

Binary erasure channel: $x \rightarrow \boxed{\text{channel}} \; y$ Goal: est $\hat{x}_1, \dots \hat{x}_n$ from $y_1, \dots y_n$

Input x_i

Output $y_i \in \{0, 1, *\}$



Perf: $P_{err} = \frac{1}{n} \sum_{i=1}^{n} P(\hat{x}_i \neq x_i)$

If we don't encode (01001)
$\Rightarrow (01*0*)$

Message len $k=5$
block len $n=5$
rate is $r = \frac{k}{n} = 1$ $\quad P_{err} = \frac{\epsilon}{2}$

If we duplicate / copy g times
rate $= \frac{1}{g}$ $\quad P_{err} = \frac{\epsilon^g}{2}$

for

Start w. a random factor graph for a LDPC $\to RG(n, m, \ell=(\ell_1, \dots \ell_n), r=(r_1, \dots, r_m))$ degree

$x_i \in \{0,1\}$ $\quad y_i \in \{0,1,*\}$ erasure

$P[y_i = * \mid x_i] = \epsilon$

$P[y_i = x_i \mid x_i] = 1-\epsilon$

$$P_y(x) = \frac{1}{Z} \underbrace{\prod_{i=1}^{n} P[y_i \mid x_i]}_{\text{Channel}} \cdot \underbrace{\prod_{a \in F} \mathbb{I}[\oplus x_a = 0]}_{\text{XOR Parity checks LDPC}}$$

$\boxed{\text{Q: What is the error rate of BPs to estimate } \cancel{x}\text{?}}$

BP on LDPC's w erasures

$$m_{i \to a}(x_i) = P[y_i \mid x_i] \prod_{b \in \partial i \setminus \{a\}} \tilde{m}_{b \to i}(x_i)$$

$$\tilde{m}_{a \to i}(x_i) = \sum_{X_{\partial a \setminus \{i\}}} \Big\{ \prod_{j \in \partial a \setminus \{i\}} m_{j \to a}(x_j) \Big\} \mathbb{I}[\oplus X_{\partial a} = 0]$$

Claim: $m_{i \to a}(x_i) \in \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} \right\} \to$ discrete

$x_{i \to a} := 1, 0, *$

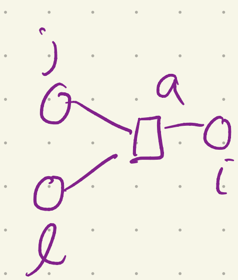
$\boxed{\text{Proof by ind}}$

init $\quad \tilde{m}^0_{b \to i}(x_i) = 1/2$

$$m^1_{i \to a}(x_i) = \begin{cases} \begin{bmatrix} \epsilon \\ \epsilon \end{bmatrix} \propto \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} & y_i = * \\ \begin{bmatrix} 1-\epsilon \\ 0 \end{bmatrix} \propto \begin{bmatrix} 1 \\ 0 \end{bmatrix} & y_i = 0 \\ \begin{bmatrix} 0 \\ 1-\epsilon \end{bmatrix} \propto \begin{bmatrix} 0 \\ 1 \end{bmatrix} & y_i = 1 \end{cases}$$

Induction

① $\tilde{m}^{(t)}_{a\to i}(x_i)$



Case 1: all incoming msgs are $\in \left\{ \begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix} \right\}$

then $\tilde{m}^{(t)}_{a\to i}(x_i) = \left[\begin{array}{l} \underline{m_{\to}(x_j=0)\, m_{\to}(x_\ell=0)} + \underline{m_{\to}(x_j=1)\, m_{\to}(x_\ell=1)} \\ \underline{m_{\to}(x_j=0)\, m(x_\ell=1)} + \underline{m_{\to}(x_j=1)\, m_{\to}(x_\ell=0)} \end{array} \right.$

$\mathbb{1}[x_i \oplus x_{\partial a \setminus \{i\}} = 0]$

exactly 1 of these 4 would be 1, others 0

Z

$= \begin{bmatrix}1\\0\end{bmatrix}$ or $\begin{bmatrix}0\\1\end{bmatrix}$

Case 2: ≥ 1 incomin msg is $\star$ / $\begin{bmatrix}1/2\\1/2\end{bmatrix}$ $\quad m_{j\to a} = \begin{bmatrix}1/2\\1/2\end{bmatrix}$

then $\tilde{m}^{(t)}_{a\to i} = \sum_{x_{j\to a}\in} \frac{1}{2}\, \mathbb{1}(x_i \oplus \underset{\substack{\| \\ \star}}{x_{j\to a}} \oplus \underset{\substack{\in \\ \{0,1\} \text{ or } \star}}{x_{\partial a\setminus\{i,j\}}}) = \begin{bmatrix}1/2\\1/2\end{bmatrix}$

then

---

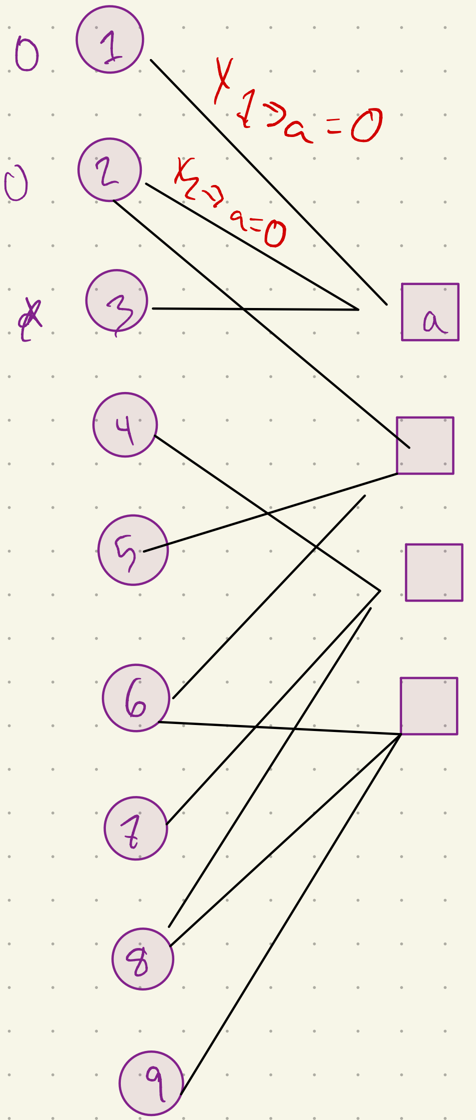
② $m^{t}_{i\to a}(x_i) = \prod_{b\in\partial i\setminus\{a\}} \tilde{m}^{t}_{b\to i}(x_i)$

(1) If all incoming $\tilde{m}_{b\to i} = \begin{bmatrix}1/2\\1/2\end{bmatrix}$, $\hat{x}_{b\to i} = \star$

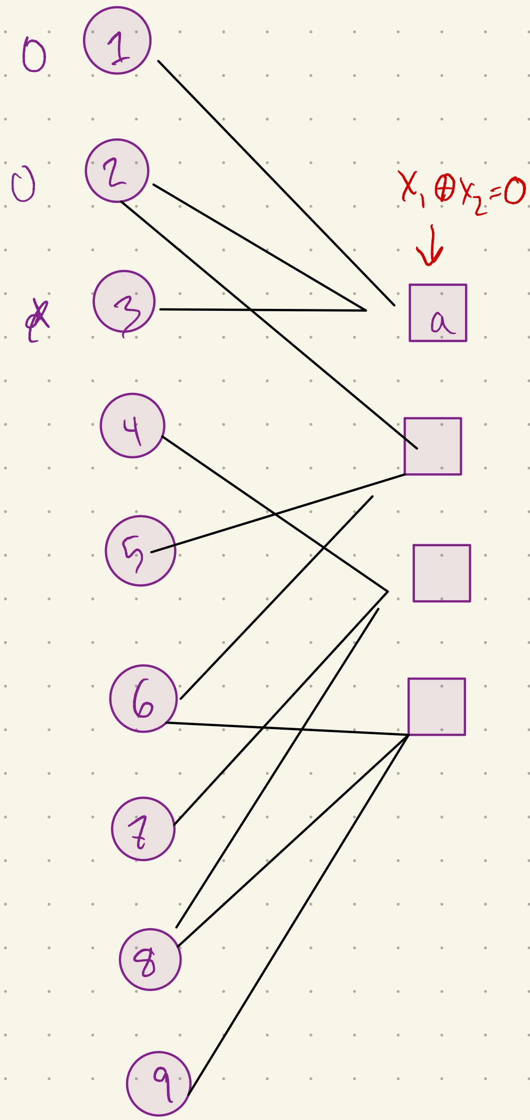
then $m_{a\to i}(x_i) = \begin{bmatrix}1/2\\1/2\end{bmatrix}$ $\quad x_{a\to i} = \star$

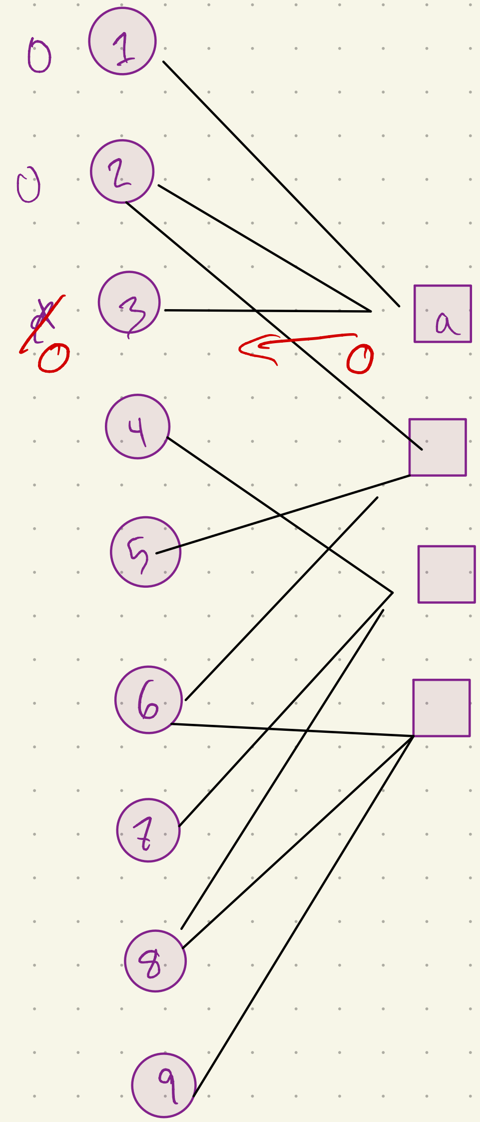
Else $m_{a\to i} \in \begin{bmatrix}0\\1\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}$

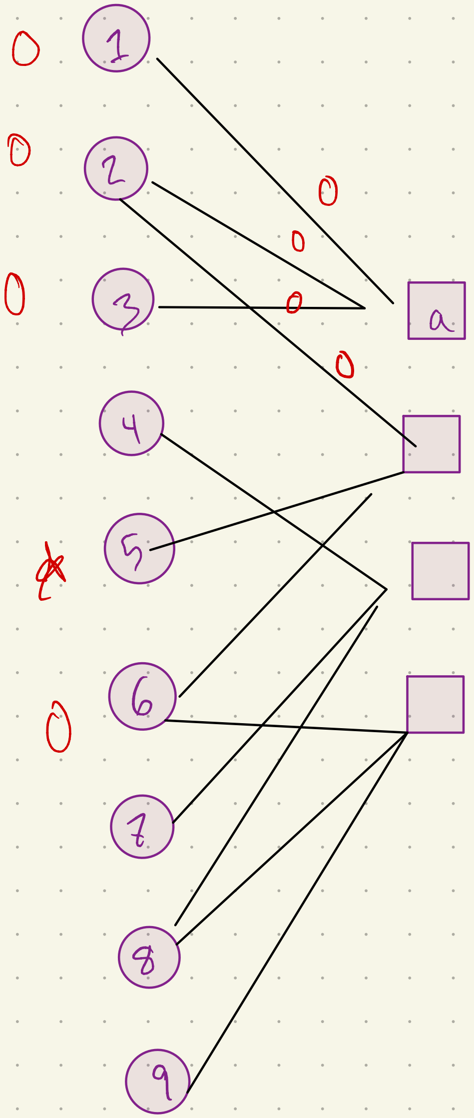
Suppose $\bar{x} = 0$



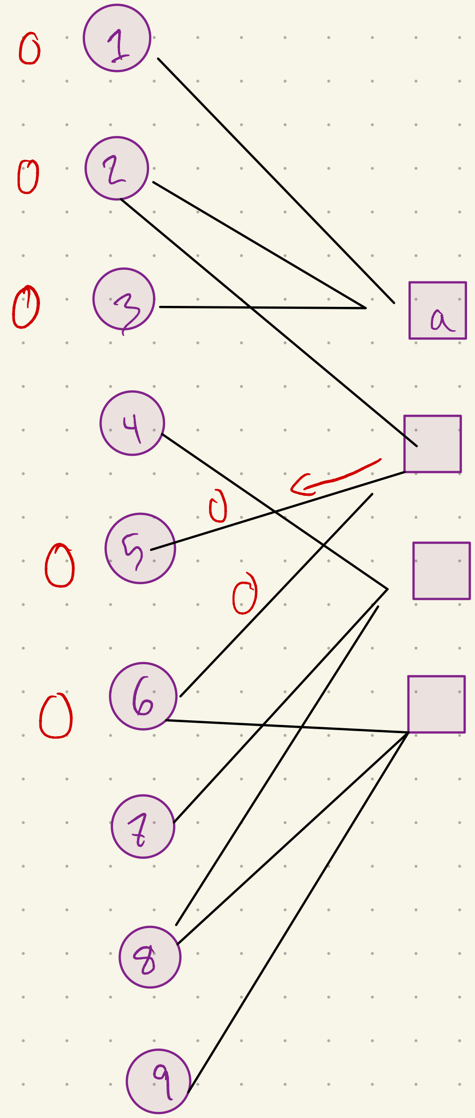
x_1 0
$x_2 = 0$







← $x_3 \oplus x_6 = 0$



If all nodes are determined / peeled, decoding success.

O/w $\exists\ S \subseteq [n]$ not resolved.

Design goal for LDPC: fixing n, m, have low $\Pr[\text{failure}]$ or low error rate $|S|$

Density evolution is critical to this design

Strategy under RFG, $m, n \to \infty$

(1) Follow BP strategy

$$\left[\begin{array}{l} \text{Update } \{m^t_{i\to a}\} \\ \text{Draw fresh RG} \\ \text{Update } \{\tilde{m}^t_{a\to i}\} \\ \text{Draw fresh RG} \\ t = t+1 \end{array}\right]$$

Analyze this process

(2) Justify it's accurate if FG is random, $m, n \to \infty$, $t \le \log n$

---

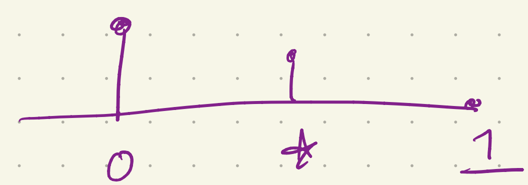
(1) Look @ computation tree for $m^{(t)}_{i\to a}(x_i)$ msg from i to a after t BP iters

$t=1$

$$\begin{pmatrix} m^1_{1\to a}(x_1) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ \vdots \\ m^1_{2\to c}(x_2) = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} \\ \vdots \end{pmatrix} \iff \begin{pmatrix} x^1_{1\to a} = 0 \\ \vdots \\ x^1_{2\to b} = * \\ \vdots \end{pmatrix}$$

$\{x^t_{i\to a}\}$

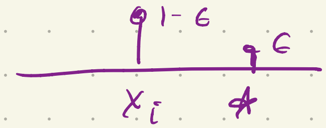
$\to$ Histogram of $\{x^t_{i\to a}\}$



Density:

$$Y_i = \begin{cases} x_i & \text{w.p. } 1-\epsilon \\ * & \text{w.p. } \epsilon \end{cases}$$

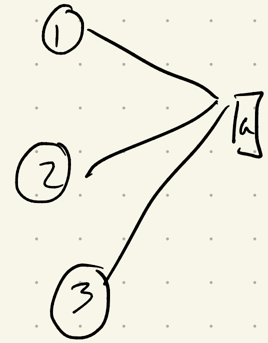
$\boxed{BEC(\epsilon)}$



$$q^{(1)} := \mathbb{P}(x_{i\to a} = *) \in [0,1] \qquad \mathbb{E}[q^{(1)}] = \epsilon$$

Density

$$\lim_{n\to\infty} q^{(1)} = \epsilon$$



$t=1$

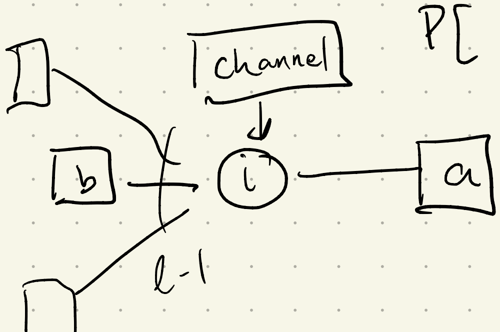
$$\tilde{x}^{(1)}_{a\to i} \in \{0,1,*\}$$

$$= \begin{cases} * & \text{if } x^{(1)}_{2\to a} = * \text{ or } x^{(1)}_{3\to a} = * \\ x_i & \text{o/w} \end{cases}$$

In general we need all $r-1$ incoming $x_{j\to a}$ to be det for $x^{(1)}_{a\to i}$ to be determined

$$\mathbb{P}[x^{1}_{j\to a} = *] = q^{(1)} \to \Pr[x^{(1)}_{a\to 1} = *] = 1-(1-q^{(1)})^{r-1} \triangleq \tilde{q}^{(1)}$$

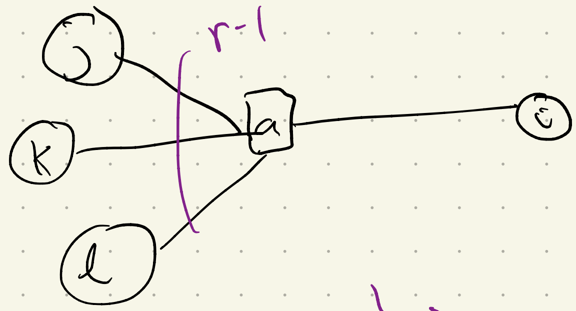
$t=2$ $\quad m^{(2)}_{i\to a}(x_i)$



$$\mathbb{P}[\tilde{x}^{(1)}_{a\to j} = *] = \tilde{q}^{(1)} \to \mathbb{P}[x^{(2)}_{i\to a} = *] = \epsilon\cdot(\tilde{q}^{1})^{\ell-1} := \tilde{q}^{(2)}$$

$$\tilde{x}_{i\to a} = \begin{cases} * & \text{if } y_i = * \ \& \text{ all } x_{b\to i} = * \\ x_i & \text{o/w} \end{cases}$$

$\tilde{m}^{(2)}_{a\to i}(x_i)$



$$\tilde{x}_{a\to i} = \begin{cases} * & \text{if } \geq 1 \text{ incoming is } * \\ \bigoplus x_{\partial a \setminus \{i\}} & \text{o/w} \end{cases}$$

$$q^{(t+1)} = \epsilon \cdot \left(\tilde{q}^{(t)}\right)^{l-1}$$

$$\tilde{q}^{(t+1)} = 1 - \left(1 - q^{t+1}\right)^{r-1} \longrightarrow q^{(t+1)} = \epsilon\left(1 - \left(1 - q^{(t)}\right)^{r-1}\right)^{l-1}$$

Depending on ϵ, l, r, $q^{(\infty)} \to 0$ [perfect decoding] or not [constant error rate]

Gaussian GMs

So far, worked with $P[X_1^n]$ $X_i \in \mathcal{X}$ $\leftarrow$ finite

- Any factor $f_a(x_1, \ldots, x_\ell)$ can be written as a $|\mathcal{X}|^\ell$-table
- Algs used only $+, *,$ look-ups

Let's start talking about <u>continuous</u> R.V.'s / dist in $\mathbb{R}^n$.

- A <u>parametric</u> <u>family</u> allows us to [store factors / compute messages

<u>Def</u> $X = (X_1, \ldots, X_n)$ is Gaussian $\mathcal{N}(\mu, \Sigma)$ if

$$P(X) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left(-\tfrac{1}{2} \cdot (x-\mu)^T \Sigma^{-1} (X-\mu)\right)$$

PDF

determinant

PD Σ

where $\mu = \mathbb{E}[X]$, $\Sigma = \mathrm{Cov}(X) = \mathbb{E}[(X-\mu)(X-\mu)^T]$