Talk on 14-15 Feb 2019

Anuran Makur

Broadcasting on Bounded Degree DAGs: [Makur-Mossel-Polyanskiy 2018]

★ Broadcasting on Trees:

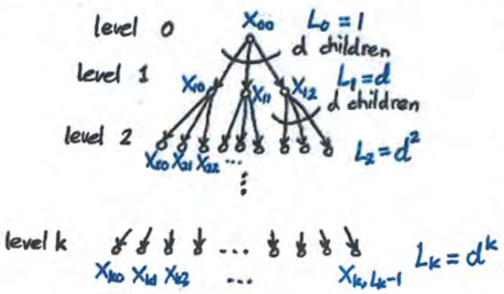
See ⑧

- Long line of work: [Bleher-Ruiz-Zagrebnov 1995] (BRZ; ↳ statistical physics result for regular trees), ..., [Evans-Kenyon-Peres-Schulman 2000] (↳ broadcasting on general trees), ... (↳ many extensions)
- Applications: phylogenetic reconstruction, random constraint satisfaction problems, etc.
- Model & Notation: (simple case of [EKPS 2000])

We are given an infinite d-ary tree. Let X_{kj} = node at the jth position in level k,

L_k = no. of nodes at level k,

$X_k = (X_{k0}, \dots, X_{k,L_k-1})$ (i.e. all nodes at level k).



Each X_{kj} is a Bernoulli random variable.

Each edge is an independent $BSC(\delta)$ with $0 < \delta < \frac{1}{2}$.

$$\{0,1\} \ni X \to \boxed{BSC(\delta)} \to Y = \begin{cases} X & \text{wp } 1-2\delta \ [\text{copy}] \\ Ber(\tfrac{1}{2}) & \text{wp } 2\delta \ [\text{ind. bit}] \end{cases} = \begin{cases} X & \text{wp } 1-\delta \\ 1-X & \text{wp } \delta \end{cases}$$

Let $X_{00} \sim Ber(\frac{1}{2})$. This defines joint distribution of $\{X_{kj} : k \geq 0, 0 \leq j < L_k\}$.

- Broadcasting Question: For what values of δ, d, Can we decode X_0 from X_k as $k \to \infty$?

$X_0 \sim Ber(\frac{1}{2}) \to X_0 = 1 : P_{X_k|X_0=1}$, $\to X_0 = 0 : P_{X_k|X_0=0}$ } Hypothesis Testing: Use ML decoder for min. prob. of error. $\hat{X}^k_{ML}(X_k)$ is the ML decoder.

- Thm: (Phase Transition)

1) If $\underbrace{(1-2\delta)^2 d > 1}_{\delta < \frac{1}{2} - \frac{1}{2\sqrt{d}}}$, then $\underbrace{\lim_{k\to\infty} \mathbb{P}(\hat{X}^k_{ML}(X_k) \neq X_0) < \frac{1}{2}}_{\triangleq \text{ broadcast/reconstruction possible}}$. [Kesten-Stigum '66]

2) If $(1-2\delta)^2 d \leq 1$, then $\lim_{k\to\infty} \mathbb{P}(\hat{X}^k_{ML}(X_k) \neq X_0) = \frac{1}{2}$. [BRZ 1995]

$\delta \geq \frac{1}{2} - \frac{1}{2\sqrt{d}}$ critical threshold

$\triangleq$ broadcast/reconstruction impossible

Intuition: (see ES section)
- $(1-2\delta)^2$ is the contraction of mutual information of a $BSC(\delta)$
- d is the "repetition code" factor
- Competition between these forces

General version uses $d = br(T) \triangleq \sup\{\lambda \geq 1 : \text{positive flow through } T \text{ s.t. edge at dist. } k \text{ has capacity } \lambda^{-k}\}$

- Observation: If $d = 1$, then broadcast impossible. So, if L_k sub-exponential, then broadcast is impossible.

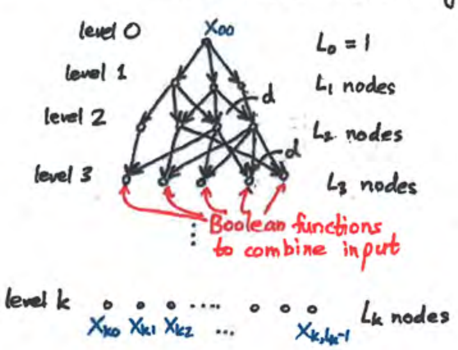
$br(T) \leq \liminf_{k\to\infty} L_k^{1/k}$

Can we have broadcast with sub-exponential L_k?

★ Existence of DAGs where Broadcasting is Possible:

- Model: (same notation as before)

We have an infinite DAG with a single source/root node in topological ordering.



As before $X_{kj} \sim$ Bernoulli and $X_{00} \sim Ber(\frac{1}{2})$, and each edge is an independent $BSC(\delta)$ with $0 < \delta < \frac{1}{2}$.

Let d = no. of incoming edges at each node. (bounded indegree)

Inputs at each node are combined using Boolean processing functions.

Question: For which δ, d, L_k, proc. function, is broadcasting possible? The processing functions allow information fusion. How small does this allow us to make L_k?

- Impossibility of Reconstruction:

Prop: If $L_k \leq \dfrac{\log(k)}{d \log(\frac{1}{2\delta})}$, then $\lim_{k\to\infty} \mathbb{P}(\hat{X}^k_{ML}(X_k) \neq X_0) = \frac{1}{2}$ regardless of our choice of processing functions.

So, the best L_k we can hope for is $L_k \geq C(\delta,d)\log(k)$ for some constant $C(\delta,d)$.

Anuran Makur

<u>Proof:</u> Let $A_k = \{$all dL_k edges from level $k-1$ to level k generate independent bits$\}$.

$\{A_k\}_{k\geq 1}$ are mutually independent and $\mathbb{P}(A_k) = (2\delta)^{dL_k}$.

$$L_k \leq \frac{\log(k)}{d\log\left(\frac{1}{2\delta}\right)} \iff (2\delta)^{dL_k} \geq \frac{1}{k}$$

$$\Rightarrow \sum_{k\geq 1} \mathbb{P}(A_k) \geq \sum_{k\geq 1} \frac{1}{k} = \infty$$

So, by Borel-Cantelli lemma, $\{A_k\}_{k\geq 1}$ occur i.o. (i.e. $\mathbb{P}(\bigcap_{m=1}^{\infty}\bigcup_{k\geq m} A_k) = 1$) almost surely.

Once an A_k occurs, all subsequent levels are independent of X_0 and the prob. of error in ML decoding $= \frac{1}{2}$. ▨

• <u>Random DAG Model:</u>

We prove <u>existence of DAGs</u> where broadcast is possible for <u>$L_k \geq C(\delta,d)\log(k)$</u> using <u>probabilistic method.</u>

DAG Model:
- Fix $L_0, L_1, L_2, \ldots$, i.e. the no. of nodes at each level, and $d \geq 3$.
- For each node $X_{k,j}$, <u>select d parents</u> in level $k-1$ <u>independently and uniformly</u> (with repetition).
- This defines a <u>random DAG G</u>. → This is strictly speaking a multigraph.

Let $\sigma_k \triangleq \frac{1}{L_k}\sum_{j=0}^{L_k-1} X_{k,j}$ be the proportion of 1's in level k.

$\{\sigma_k : k\in\mathbb{N}\}$ forms a <u>Markov chain</u>, and <u>σ_k</u> is a <u>sufficient statistic of X_k for</u> performing inference about $\underset{\overset{\|}{X_0}}{\sigma_0}$.

↳ <u>Intuition:</u> order of $X_{k,j}$ in X_k does not matter. (<u>Fisher-Neyman:</u> $P_{X_k|\sigma_0} = h(X_k)\, g(\sigma_k,\sigma_0)$)

$$\delta_{maj} \triangleq \frac{1}{2} - \frac{2^{d-2}}{\lceil\frac{d}{2}\rceil\binom{d}{\lceil\frac{d}{2}\rceil}} \approx \frac{1}{2}\left(1 - \frac{\sqrt{\pi}}{\sqrt{d}}\right)$$

• <u>Thm:</u> Let δ_{maj} & all processing functions be <u>majority</u> (ties broken randomly).

1) If $0 < \delta < \delta_{maj}$ and $L_k \geq C(\delta,d)\log(k)$ for some $C(\delta,d)$, then $\limsup_{k\to\infty} \mathbb{P}\big(\underbrace{\mathbb{1}\{\sigma_k \geq \tfrac{1}{2}\}}_{\text{majority decoder}} \neq \sigma_0\big) < \frac{1}{2}$. (⇒ broadcast possible with ML decoder)

2) If $\underbrace{\delta_{maj}}_{\text{critical threshold}} < \delta < \frac{1}{2}$ and $L_k = o(D(\delta,d)^{-k})$, then $\lim_{k\to\infty} \| P^+_{X_k|G} - P^-_{X_k|G} \|_{TV} = 0$ G-a.s. (⇔ $\mathbb{P}(\hat{X}^k_{ML}(X_k,G) \neq X_0 | G) = \frac{1}{2}\left(1 - \|P^+_{X_k|G} - P^-_{X_k|G}\|_{TV}\right) \xrightarrow{k\to\infty} \frac{1}{2}$ G-a.s. i.e. broadcast impossible a.s. with ML decoder that knows G)

Lip. const $= D(\delta,d) < 1$ over here

$P^+_{X_k|G} = P_{X_k|G,X_0=1}$, $\quad P^-_{X_k|G} = P_{X_k|G,X_0=0}$

• <u>Intuition:</u> (<u>$d=3$ case:</u> $\delta_{maj} = \frac{1}{6}$)

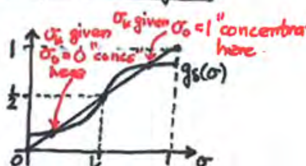
Given $\sigma_{k-1} = \sigma$, $X_{k,j} \overset{iid}{\sim} \mathrm{maj}(\mathrm{Ber}(\sigma*\delta), \mathrm{Ber}(\sigma*\delta), \mathrm{Ber}(\sigma*\delta))$. ← indep. Ber. vars; $\sigma*\delta = \sigma(1-\delta)+\delta(1-\sigma)$

$$\Rightarrow \mathbb{P}(X_{k,j}=1 | \sigma_{k-1}=\sigma) = (\sigma*\delta)^3 + 3(\sigma*\delta)^2(1-\sigma*\delta) \triangleq g_\delta(\sigma) \leftarrow \text{cubic poly. in } \sigma$$

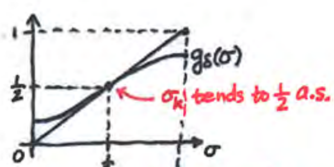
Since $L_k\sigma_k \sim \mathrm{binomial}(L_k, g_\delta(\sigma)) \,|\, \sigma_{k-1}=\sigma$, <u>$\mathbb{E}[\sigma_k|\sigma_{k-1}=\sigma] = g_\delta(\sigma)$</u>.

For large k, given $\sigma_{k-1}=\sigma$, $\sigma_k \approx \mathbb{E}[\sigma_k|\sigma_{k-1}=\sigma] = g_\delta(\sigma)$.

<u>Fixed point analysis:</u>



<u>Case:</u> $\delta < \frac{1}{6}$ (3 fixed points)



<u>Case:</u> $\delta \geq \frac{1}{6}$ (1 fixed point)

We require concentration inequalities to rigorize this intuition.

• <u>Proof of Part 2:</u> ($\delta > \delta_{maj}$) ← general $d\geq 3$

First, construct <u>monotone coupling</u> of $\{X^+_k : k\in\mathbb{N}\}$ ($\{X_k\}$ started at $X_0=1$) and $\{X^-_k : k\in\mathbb{N}\}$ ($\{X_k\}$ started at $X_0=0$). So, we have $\{(X^+_k, X^-_k) : k\in\mathbb{N}\}$ s.t. <u>$X^+_{k,j} \geq X^-_{k,j}$ a.s., $\forall k,j$</u>. ← Markovian coupling

$$\|P^+_{X_k|G} - P^-_{X_k|G}\|_{TV} \underset{\text{Dobrushin}}{\leq} \mathbb{P}(X^+_k \neq X^-_k|G) \leq \sum_{j\in[L_k]} \mathbb{P}(\underbrace{X^+_{k,j} - X^-_{k,j}}_{\geq 0 \text{ a.s.}} \neq 0|G) = L_k\,\mathbb{E}\Big[\sum_j \tfrac{1}{L_k}(X^+_{k,j} - X^-_{k,j})\Big|G\Big] = L_k\,\mathbb{E}[\underbrace{\sigma^+_k - \sigma^-_k}_{\geq 0\text{ a.s.}}|G] \Rightarrow \mathbb{E}\big[\|P^+_{X_k|G} - P^-_{X_k|G}\|_{TV}\big] \leq L_k\,\mathbb{E}[\sigma^+_k - \sigma^-_k].$$

$$\mathbb{E}[\sigma^+_k - \sigma^-_k] = \mathbb{E}\big[\mathbb{E}[\sigma^+_k - \sigma^-_k | \sigma^+_{k-1}, \sigma^-_{k-1}]\big] = \mathbb{E}\big[\underbrace{g_\delta(\sigma^+_{k-1}) - g_\delta(\sigma^-_{k-1})}_{\geq 0\text{ a.s.}}\big] \leq \underbrace{D(\delta,d)}_{\text{Lip. constant} < 1 \text{ for } \delta > \delta_{maj}} \mathbb{E}[\sigma^+_{k-1} - \sigma^-_{k-1}] \overset{\text{recursion}}{\leq} (D(\delta,d))^k\, \underbrace{\mathbb{E}[\sigma^+_0 - \sigma^-_0]}_{=1} = D(\delta,d)^k$$

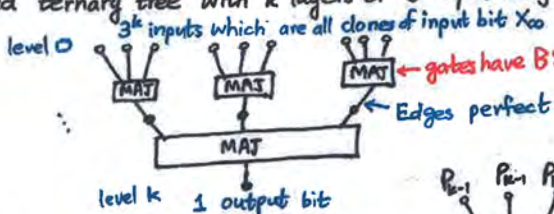
$$\mathbb{E}\big[\|P^+_{X_k|G} - P^-_{X_k|G}\|_{TV}\big] \leq L_k\,\mathbb{E}[\sigma^+_k - \sigma^-_k] \leq L_k D(\delta,d)^k \xrightarrow[k\to\infty]{\geq 0 \text{ a.s.}} 0 \text{ when } L_k = o(D(\delta,d)^{-k}). \quad ▨$$

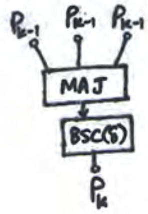
[<u>Note:</u> If $\sum_{k=0}^{\infty} L_k D(\delta,d)^k < \infty$, then $\forall\varepsilon>0$, $\sum_{k=1}^{\infty} \mathbb{P}(\|P^+_{X_k|G} - P^-_{X_k|G}\|_{TV} > \varepsilon) \leq \frac{1}{\varepsilon}\sum_{k=1}^{\infty}\mathbb{E}\big[\|P^+_{X_k|G} - P^-_{X_k|G}\|_{TV}\big] < \infty \overset{\text{B-C Lemma}}{\Rightarrow} \lim_{k\to\infty} \|P^+_{X_k|G} - P^-_{X_k|G}\|_{TV} = 0$ G-a.s.]

• **Remarks:**

① Von Neumann Model of Storing a Bit: [Von Neumann 1956] (cf. [Hajek-Weller 1991] – $d=3$ case, [Evans-Schulman 2003] – general case)

Balanced ternary tree with k layers of 3-input majority gates.



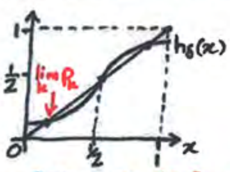


Consider a single noisy majority gate:
Let P_k = prob. of error (i.e. value at node $\neq$ value of input)

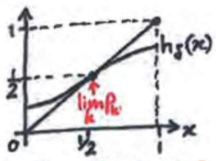
recursion $\begin{cases} P_k = \underbrace{\delta}_{\text{generate wrong bit indep'ly in BSC}} + \underbrace{(1-2\delta)}_{\text{copy in BSC}}\underbrace{(3P_{k-1}^2 - 2P_{k-1}^3)}_{\text{prob. at least 2 inputs of MAJ are wrong (all bits at input are supposed to be equal)}} \\ P_0 = 0 \end{cases}$

Define $h_\delta(x) = \delta + (1-2\delta)(3x^2 - 2x^3)$. Then, $P_k = \underbrace{h_\delta^{(k)}}_{\text{k-fold composition}}(0)$. $\boxed{h_\delta^{(k+1)}(0) = \delta * g_\delta^{(k)}(0)}$ } induction

Fixed Point Analysis:



Case: $\delta < \frac{1}{6}$ (3 fixed points)



Case: $\delta \geq \frac{1}{6}$ (1 fixed point)
↑ THRESHOLD $\frac{1}{6}$ is the SAME !

$$\boxed{\begin{array}{l} \text{Hence, if } \delta < \frac{1}{6}\text{, then } \lim_{k\to\infty} P_k < \frac{1}{2} \text{ (storage possible)}, \\ \text{\& if } \delta \geq \frac{1}{6}\text{, then } \lim_{k\to\infty} P_k = \frac{1}{2} \text{ (storage impossible)}. \end{array}}$$

② Comparison to our model:

• We need concentration of measure on top of the fixed point analysis.

③ Open Questions: If $d=3$ and $L_k = O(\log(k))$ (random DAG model), reconstruction is impossible for all choices of Boolean processing functions when $\delta \geq \frac{1}{6}$. (Equivalently, majority processing functions are "optimal.")

④ Evans-Schulman Estimate: [Evans-Schulman 1999] For deterministic DAGs, $I(X_0;X_k) \leq L_k((1-2\delta)^2 d)^k$.
Hence, if $L_k = o\left(\left(\frac{1}{(1-2\delta)^2 d}\right)^k\right)$ and $\underbrace{(1-2\delta)^2 d < 1}_{\delta > \frac{1}{2} - \frac{1}{2\sqrt{d}}}$, then $\lim_{k\to\infty} I(X_0;X_k) = 0 \underset{\text{Pinsker}}{\Rightarrow} \lim_{k\to\infty} \|P_{X_k}^+ - P_{X_k}^-\|_{TV} = 0$.

$\boxed{d=3 \text{ case}}$ { This means $(1-2\delta)^2 3 < 1$ ($\Leftrightarrow \delta > \frac{1}{2} - \frac{1}{2\sqrt{3}} = 0.211\ldots$) implies that reconstruction is impossible for all deterministic DAGs for any choice of processing functions.
↳ as well as random DAGs

• Cor: (Existence) For any $d \geq 3$, $\delta < \delta_{maj}$, there exists a (deterministic) DAG with $L_k \geq C(\delta,d)\log(k)$ and majority processing functions such that $\lim_{k\to\infty} \mathbb{P}(\hat{X}_{ML}^k(X_k) \neq X_0) < \frac{1}{2}$.
↑ ML decoder with knowledge of DAG

Proof: Fix $d \geq 3$, $L_k \geq C\log(k)$, and $\delta < \delta_{maj}$.
From theorem, $\exists \varepsilon > 0$ (deps. on δ, d) s.t. $\mathbb{P}(\mathbb{1}\{\sigma_k \geq \frac{1}{2}\} \neq X_0) \leq \frac{1}{2} - 2\varepsilon$ for all suff. large k.
Let $P_k(G) \triangleq \mathbb{P}(\hat{X}_{ML}^k(X_k|G) \neq X_0 | G)$ be the prob. of error in ML decoding given G (random DAG realization).
↑ ML decoder knows G
Let $E_k = \{\text{all DAGs } \mathcal{G} \text{ s.t. } P_k(\mathcal{G}) \leq \frac{1}{2} - \varepsilon\}$.

$\frac{1}{2} - 2\varepsilon \geq \mathbb{P}(\mathbb{1}\{\sigma_k \geq \frac{1}{2}\} \neq X_0) \underset{\text{ML decoder optimal}}{\geq} \mathbb{E}[P_k(G)] \geq \mathbb{E}[P_k(G)|G \notin E_k]\mathbb{P}(G \notin E_k) \overset{\text{def. of } E_k}{\geq} (\frac{1}{2} - \varepsilon)\mathbb{P}(G \notin E_k) \Rightarrow \mathbb{P}(G \in E_k) \geq \frac{2\varepsilon}{1-2\varepsilon} > 0$ for all suff. large k

Since $\{E_k\}_{k\geq 1}$ is a decreasing sequence (as $P_k(\cdot)$ is increasing in k), $\mathbb{P}(G \in \bigcap_{k\geq 1} E_k) = \lim_{k\to\infty} \mathbb{P}(G \in E_k) \geq \frac{2\varepsilon}{1-2\varepsilon} > 0$. ▨

Anuran Makur

- What about $d=2$?

Thm: Let $d=2$, all processing functions at even levels be AND, all processing functions at odd levels be OR. and $L_k \geq C(\delta)\log(k)$ ← different from $C(\delta)$ earlier

random DAG model

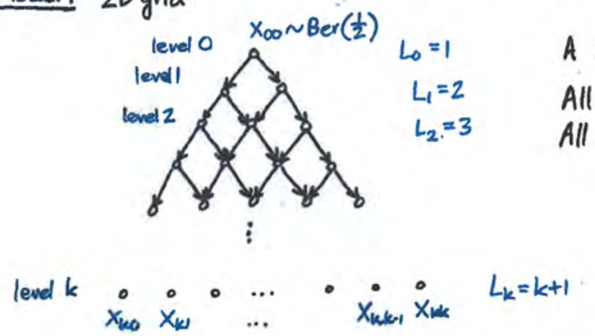
1) If $0<\delta<\frac{3-\sqrt{7}}{4}=0.088\ldots$ ∧ and $L_k = o(E(\delta)^{-k})$, then $\limsup_{k\to\infty} \mathbb{P}\big(\underbrace{\mathbb{1}\{\sigma_{2k}\geq t(\delta)\}}_{\text{biased majority decoder}} \neq \sigma_0\big) < \frac{1}{2}$.

2) If $\underbrace{\frac{3-\sqrt{7}}{4}}_{\text{critical threshold}}<\delta<\frac{1}{2}$ ∧, then $\lim_{k\to\infty} \|P^+_{2k|G}-P^-_{2k|G}\|_{TV}=0$ G-a.s.

Threshold in [Evans-Pippenger '98] and [Unger 2007]

★ Regular 2D Grids – Impossibility of Broadcasting:

- Model: 2D grid



A 2D grid is a specific DAG (deterministic).
All processing functions with 2 inputs are the same.
All processing functions with 1 input are the identity.

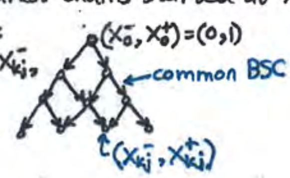
- Conjecture: Broadcasting is impossible for 2D grids (as defined above) regardless of the noise level δ.
- Why?
  ① Random DAG view ⇒ one fixed point for all $\delta\in(0,\frac{1}{2})$ when $d=2$, $L_k=k+1$. (naïve, possibly misleading)
  ↳ Proof of deterministic case is much more difficult.

  ② PCA view ⇒ If reconstruction is possible in 2D grid, then 2D grid is "not ergodic."
  This suggests existence of 1D PCA with binary state space that is not ergodic.
  (Known constructions require a lot more states [Gács 2001].)
  So, 2D grid should be "ergodic."

  Note: PCA different from 2D grid because:
  a) PCA uses weak convergence, while we use TV convergence,
  b) 2D grid is a PCA with boundary conditions. (PCA has stronger separation between all zeros and all ones initial configs.)

- **Thm:** If all processing functions are AND, or all are XOR, then $\underbrace{\lim_{k\to\infty}\|P_{X_k}^+ - P_{X_k}^-\|_{TV}=0}_{\text{broadcast impossible}}$ regardless of $\delta\in(0,\frac{1}{2})$.

- Proof of AND case: (Sketch)
  ① Monotone Markovian Coupling:
  Let $\{X_k^+ : k\in\mathbb{N}\}$ and $\{X_k^- : k\in\mathbb{N}\}$ be the Markov chains started at $X_0^+=1$ and $X_0^-=0$, respectively.
  We "run" these chains on the same grid:
  → Each BSC(δ) either copies both X_{kj}^+ and X_{kj}^-, or generates the same independent bit for both chains.



  Then, $\forall k,j$, $X_{kj}^+ \geq X_{kj}^-$ a.s.. [Check this]

[continued.]

Proof of AND case cont'd:

Let each node of the grid be $Y_{kj} \triangleq (X_{kj}^-, X_{kj}^+) \in \{\underset{(0,0)}{0_c}, \overset{(0,1)}{1_u}, \underset{(1,1)}{1_c}\}$. ← (1,0) is not required in the alphabet

(BSC has matrix $W = \begin{array}{c|ccc} & 0_c & 1_u & 1_c \\ \hline 0_c & 1-\delta & 0 & \delta \\ 1_u & \delta & 1-2\delta & \delta \\ 1_c & \delta & 0 & 1-\delta \end{array}$, and AND operates entrywise.)

② Reduction to Coupled Grid:

$\|P_{X_k}^+ - P_{X_k}^-\|_{TV} \underset{\text{Dobrushin}}{\leq} \mathbb{P}(X_k^+ \neq X_k^-) = 1 - \mathbb{P}(X_k^+ = X_k^-)$

Since $(\{X_k^+ = X_k^-\})_{k\geq 1}$ is increasing, $\lim_{k\to\infty} \|P_{X_k}^+ - P_{X_k}^-\|_{TV} \leq 1 - \mathbb{P}(\exists k, X_k^+ = X_k^-)$.

So, it suffices to prove that: $\mathbb{P}(A) = 1$ for $A \triangleq \{\exists k$, there are no 1_u's in level $k\} = \{\exists k, X_k^+ = X_k^-\}$.

③ Oriented Bond Percolation: [Durrett 1984]

Remove each edge indep'ly wp $1-p$, ("closed")
or keep it wp $p \in [0,1]$. ("open")

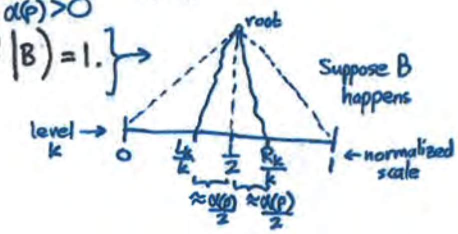
Let $B = \{\exists$ infinite open path starting at root$\}$,
R_k = rightmost node index at level k that is connected to root,
L_k = leftmost ——"—— .

eg:

Thm: For phase transition threshold $\delta_{perc} \in (\frac{1}{2}, 1)$:

1) If $p > \delta_{perc}$, then $\mathbb{P}_p(B) > 0$ and $\mathbb{P}_p\left(\lim_{k\to\infty} \frac{R_k}{k} = \frac{1+\alpha(p)}{2} \text{ and } \lim_{k\to\infty}\frac{L_k}{k} = \frac{1-\alpha(p)}{2} \,\middle|\, B\right) = 1.$ ← measure depends on p; for some $\alpha(p) > 0$

2) If $p < \delta_{perc}$, then $\mathbb{P}_p(B) = 0$.



④ Case I: $p = 1-2\delta < \delta_{perc}$ $\left(\Leftrightarrow \delta > \frac{1-\delta_{perc}}{2}\right)$

Edge open ⇔ BSC copies.

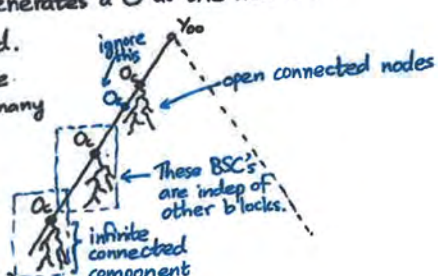
By Thm part 2, $\mathbb{P}(B) = 0 \Leftrightarrow \mathbb{P}(\{$no. of nodes connected to root with copies is finite$\}) = 1$. (B^c)

Since B^c occurs a.s., there is a level where no BSCs copy 1_u's ⇒ there is a level k with no 1_u's. ($B^c \subseteq A$)

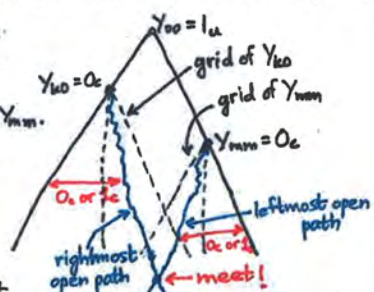
Hence, $\mathbb{P}(A) = 1$.

⑤ Case II: $p = 1-\delta > \delta_{perc}$ $(\Leftrightarrow \delta < 1 - \delta_{perc})$

- Edge open ⇔ BSC copies or generates a 0 as the new bit.
- Consider "left edge" of 2D grid.
- Since BSC's on this side generate indep 0 wp δ, there are infinitely many $Y_{k0} = 0_c$ on the left side.
- Each 0_c has a set of open connected nodes below it, and the set is ∞ wp $\mathbb{P}(B) > 0$. — Use Thm part 1
- Since blocks of (0_c, connected nodes) are independent, we almost surely have $Y_{k0} = 0_c$ with an infinite open path connected to it. (Borel-Cantelli)



- Similarly, $\exists m$ s.t. $Y_{mm} = 0_c$ with an infinite open path connected to it a.s..
- By Thm part 1, the rightmost path from Y_{k0} meets the leftmost path from Y_{mm}.
- All nodes on these paths are 0_c, and all nodes enclosed by these paths are 0_c or 1_c!



∴ When the paths meet, all nodes at that level are not 1_u's. Hence, $\mathbb{P}(A) = 1$. ∎

Anuran Makur

★ Miscellaneous Notes:

1 Evans-Schulman Estimate: [Evans-Schulman 1999] (c.f. [Polyanskiy-Wu 2017])

Consider a Bayesian network on a DAG with one source node X.
For any node W, we identify W with the random variable at W.
Let $\underbrace{\eta_W}_{\text{contraction coefficient}} \triangleq \eta_{KL}(P_{W|\underbrace{pa(W)}_{\text{parents of } W}})$, and for any path $\pi = (\underbrace{v_0}_{\text{start}}, \dots, \underbrace{v_k}_{\text{end}})$, $\eta_\pi \triangleq \prod_{i=1}^{k} \eta_{v_i}$.

Note: For channel $P_{Y|X}$, $\eta_{KL}(P_{Y|X}) \triangleq \sup_{\substack{P_{U,X}:\\ U\to X\to Y}} \dfrac{I(U;Y)}{I(U;X)}$.

→ **Thm:** $\eta_{KL}(P_{V|X}) \leq \sum_{\pi: X \to V} \eta_\pi$ for every set of nodes V. } path from X to V

Proof: Order all nodes in the DAG (so that $\underbrace{\text{ord}}_{\text{index in order}}(X) = 0$) and the ordering is consistent with the topological ordering. For a set of nodes V, let $\text{ord}(V) = \sup\{\text{ord}(W) : W \in V\}$.

Suppose $W > V$ for a node W and set of nodes V, i.e. $\text{ord}(W) > \text{ord}(V)$.

$\boxed{\text{Claim: } \eta_{KL}(P_{W,V|X}) \leq \eta_W\, \eta_{KL}(P_{V,pa(W)|X}) + (1-\eta_W)\, \eta_{KL}(P_{V|X}).}$

Pf: Consider Markov chain $U \to X \to (V,A) \to W$ for arbitrary U and $A = pa(W) - V$.
Given V, we still have $U \to X \to A \to W$.
Hence, $I(U;W|V=v) \leq \underbrace{\eta_{KL}(P_{W|A,V=v})}_{\leq\, \eta_{KL}(P_{W|pa(W)}) \text{ [by definition]}} I(U;A|V=v) \Rightarrow I(U;W|V) \leq \eta_W\, I(U;A|V)$.

Adding $I(U;V)$ to both sides gives:

$$I(U;W,V) \leq \eta_W\, I(U;V,A) + (1-\eta_W)\, I(U;V)$$

$$\Rightarrow \frac{I(U;W,V)}{I(U;X)} \leq \eta_W \frac{I(U;V,A)}{I(U;X)} + (1-\eta_W)\frac{I(U;V)}{I(U;X)}$$

$$\Rightarrow \eta_{KL}(P_{W,V|X}) \leq \eta_W\, \eta_{KL}(P_{V,pa(W)|X}) + (1-\eta_W)\, \eta_{KL}(P_{V|X}). \quad \blacksquare$$

The rest of proof follows by strong induction on the ord(·) of sets of nodes.

Inductive hypothesis → Assume $\eta_{KL}(P_{V|X}) \leq \sum_{\pi: X\to V} \eta_\pi$ for every set of nodes V with $\text{ord}(V) \leq k$.

Let $\text{ord}(W) = k+1$ for node W. Then for any V with $\text{ord}(V) \leq k$, we have:

$$\underbrace{\eta_{KL}(P_{W,V|X})}_{\text{ord}(W,V)=k+1} \leq \eta_W\, \underbrace{\eta_{KL}(P_{V,pa(W)|X})}_{\text{ord}(V,pa(W)) \leq k} + (1-\eta_W)\,\eta_{KL}(P_{V|X})$$

$$\leq \eta_W \sum_{\pi: X\to V, pa(W)} \eta_\pi + (1-\eta_W) \sum_{\pi: X\to V} \eta_\pi$$

$$= \eta_W \sum_{\pi: X\to A} \eta_\pi + \sum_{\pi: X\to V} \eta_\pi$$

$$\leq \eta_W \sum_{\pi: X\to pa(W)} \eta_\pi + \sum_{\pi: X\to V} \eta_\pi \quad [\text{as } pa(W) \supseteq A]$$

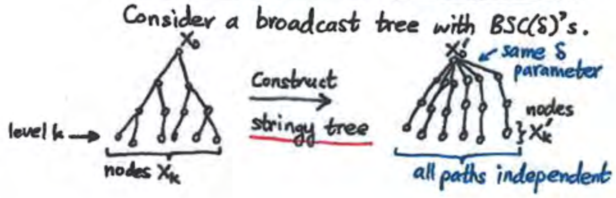
$$= \sum_{\pi: X\to W} \eta_\pi + \sum_{\pi: X\to V} \eta_\pi$$

$$= \sum_{\pi: X\to W,V} \eta_\pi .$$

Hence, the result is true for all sets of nodes V with $\text{ord}(V) \leq k+1$.
The proof is complete by induction. $\blacksquare$

2 Evans-Schulman Estimate for Trees: [Evans-Kenyon-Peres-Schulman 2000]

Consider a broadcast tree with BSC(δ)'s.



$\eta_{KL}(BSC(\delta)) = (1-2\delta)^2$ [Ahlswede-Gács '76]

Thm: $\underbrace{I(X_0;X_k) \leq I(X_0';X_k')}_{\text{degradation construction!}} \leq \underbrace{\sum_i I(X_0';X_{ki}')}_{X_{ki}' \text{ cond. iid given } X_0'} \leq \underbrace{d^k}_{\text{no. of paths}} \underbrace{(1-2\delta)^{2k}}_{\text{contraction on path}} = ((1-2\delta)^2 d)^k$.

So, if $(1-2\delta)^2 d < 1$, reconstruction is impossible.

$X_0 \to X_k' \to X_k$, $X_0 = X_0'$

Degradation $\{ P_{X_k|X_0} = \underbrace{P_{X_k'|X_0'}}_{\text{row stoch.}} \cdot \underbrace{\tilde{P}_{X_k|X_k'}}_{\text{some channel}}$

Anuran Makur

③ Probabilistic Cellular Automata: (1D) ← abbrev. as PCA

- sites $\mathbb{Z}$
- state space S, $|S|<\infty$ (eg: $S=\{0,1\}$)
- configuration space $S^{\mathbb{Z}}$ (functions $\xi:\mathbb{Z}\to S$)
- deterministic function $f: S^{|\mathcal{N}|}\to S$ (e.g. $f=\text{MAJ}$)
- neighborhood $\mathcal{N}$ (eg: $\{-1,0,1\}$) ← $|\mathcal{N}|<+\infty$

eg:

↑ At time $k+1$, each site x simultaneously computes $f(\xi_k(x+i): i\in\mathcal{N})$ and passes it through an indep. BSC(δ) to get $\xi_{k+1}(x)$. } Dynamics

Main Question: Is a PCA ergodic?

PCA defines a Markov process on $S^{\mathbb{Z}}$. For any initial config. $\xi\in S^{\mathbb{Z}}$, let ν_k^{ξ} be the prob. measure on $S^{\mathbb{Z}}$ at time k.

We say a PCA is ergodic $\Leftrightarrow$ $\exists!$ invariant measure ν_0 on $S^{\mathbb{Z}}$ s.t. $\forall$ init. config. $\xi\in S^{\mathbb{Z}}$, $\nu_k^{\xi}\xrightarrow{w}\nu_0$ as $k\to\infty$. ← weak convergence

Weak Convergence:

For $S^{\mathbb{Z}}$, the σ-algebra is defined as follows.

$$\mathcal{C}\triangleq\bigcup_{\substack{A\subseteq\mathbb{Z},\\|A|<\infty}}\left\{\{\xi\in S^{\mathbb{Z}}:\xi(A)=x_A\}\,\middle|\,x_A\in S^{|A|}\right\}$$ ← cylinder sets

$\sigma(\mathcal{C})$ is the σ-algebra on $S^{\mathbb{Z}}$.

By Daniell-Kolmogorov theorem, consistent finite-dim marginals defined on $\mathcal{C}$ uniquely determine measure on $S^{\mathbb{Z}}$.

Def: μ_n,μ measures on $(S^{\mathbb{Z}},\sigma(\mathcal{C}))$.
$\mu_n\xrightarrow{w}\mu \Leftrightarrow \forall C\in\mathcal{C}, \lim_{n\to\infty}\mu_n(C)=\mu(C)$. ← weak convergence corresponds to convergence of finite-dim distributions

Note: $\mu_n\xrightarrow{w}\mu \Leftrightarrow \forall C\in\mathcal{C}, \lim_{n\to\infty}\mu_n(C)=\mu(C)$

$\not\Leftrightarrow \lim_{n\to\infty}\sup_{A\in\sigma(\mathcal{C})}|\mu_n(A)-\mu(A)| \Leftrightarrow \lim_{n\to\infty}\|\mu_n-\mu\|_{TV}$

i.e. weak convergence $\not\Rightarrow$ TV convergence.

For many PCA with special characteristics, ergodicity can be determined by convergence of dist.s over finite intervals (e.g. single sites).

"Positive Rates" Conjecture: A 1D PCA ← discrete-time (could be cont. too) with $S=\{0,1\}$ with:

simple case where $S=\{0,1\}$ and noise dist. is unif & error rate is 2δ {
1. finite $\mathcal{N}$, i.e. $|\mathcal{N}|<\infty$
2. strictly positive δ (>0) ← positive rates condition
}

must be ergodic.

[Gács 2001] gave counter-example, but in simple settings such as the one above, it is still open.

★ Motivation:

① **Ising Models:** $G = (V, E)$ Let $\sigma_v \in \{\pm 1\}$ be spin at $v \in V$.

$$\underbrace{P(\{\sigma_v : v \in V\})}_{\text{Gibbs dist.}} \triangleq \underbrace{Z(t)^{-1}}_{\text{partition function}} \exp\Big(\overbrace{\sum_{(u,v)\in E} \underbrace{J}_{\text{interaction potential}>0\ \text{(ferromagnetic)}} \sigma_u \sigma_v / \underbrace{t}_{\text{temperature}>0}}^{\text{Hamiltonian}}\Big)$$

Relation to broadcasting:

$$\frac{\delta}{1-\delta} = \exp(-2J/t) \quad \text{for } G = \text{tree}$$

$(\delta \to \tfrac{1}{2}$ as $t \to \infty)$

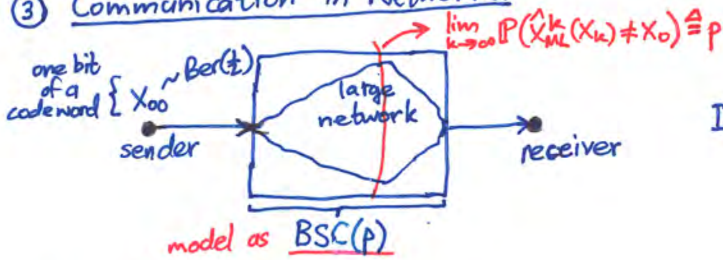
Define limiting Gibbs states using DLR boundary conditions, and let $\mathcal{G}$ (includes all possible boundary conditions) be convex set of Gibbs states (which is non-empty). The limiting Gibbs state μ with free boundary conditions (which is our model) is extremal in $\mathcal{G}$ (i.e. not a convex combination of measures in $\mathcal{G}$) iff broadcasting impossible.

↳ $\mu = \frac{1}{2}\mu^+ + \frac{1}{2}\mu^-$ always, but $\mu^+ = \mu^-$

(Note: Broadcasting possible $\Rightarrow \mu$ is $\frac{1}{2}\underbrace{\mu^+}_{} + \frac{1}{2}\underbrace{\mu^-}_{}$.)
initial conditions at root
$\mu^+ \neq \mu^-$

② **Genetic Reconstruction:** (also phylogenetic tree reconstruction)
Reconstruct traits of ancestors from observed population.
Preffered method is parsimony, which is equivalent to ML for small δ.

③ **Communication in Networks:**



If broadcasting impossible $\Rightarrow$ coding impossible.

④ Community Detection
⑤ Random Constraint Satisfaction
⑥ PCA (see ⑦)
⑦ **Reliable Computation/Storage:** [von Neumann '56], [Hajek-Weller '91], [Evans-Schulman '03], [Unger '07]
Our threshold matches threshold of reliable computation using formulae.
(Thm: For odd $d \geq 3$, δ, $\exists\, \varepsilon = \varepsilon(\delta, d) \in (0, \frac{1}{2})$ s.t. all Boolean functions can be computed using δ-noisy formulae (with d-input gates) for all inputs with $\mathbb{P}[\text{error}] \leq \frac{1}{2} - \varepsilon$ iff $\delta < \delta_{maj}$.)
