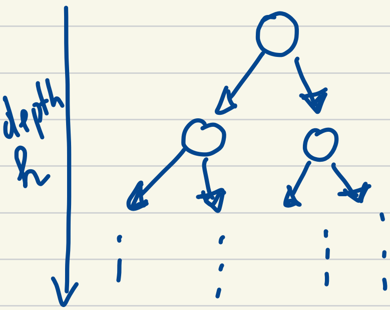
6.S896 - Algorithmic Statistics

Lecture 9: The role of Temperature (part 1)

↳ Kesten-Stigum Bound

• Setup: Broadcasting Model on a Tree



- complete binary tree T of depth h
  ↳ things appropriately extend to other degrees & irregular trees
  ↳ root node ρ
  ↳ edges directed away from ρ

- Random Process: samples r.v. $X_v \in \{\pm 1\}$ at each $v \in V$

  • $X_\rho \overset{u.a.r}{\sim} \{\pm 1\}$
  • processing all other nodes v following a topological ordering of T, sample X_v conditioning on X_u as follows

$$\Pr[X_v = s \mid X_u = s] = 1 - p \quad , \forall s \in \{\pm 1\}$$

aka $\Pr[X_v = -s \mid X_u = s] = p$
↳ "mutation probability"

assume $p \in [0, \frac{1}{2}]$ makes exposition slightly easier (& is w.l.o.g. up to flipping $1 \leftrightarrow -1$ every other layer)

↳ thus each edge: $u \to v$ is a binary symmetric channel $M = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}$

↳ things extend to other channels

- Very convenient decomposition of M

$$\begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix} = \underbrace{(1-2p)}_{\lambda_2(M)}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + 2p\begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$$

↳ second largest eigenvalue of M corresponding to eigenvector $\begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$ (the other eigenvalue is 1 corresponding to $\begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$)

↳ interpretation: with probability $\lambda = 1-2p$: $X_v = X_u$

with prob. $1-\lambda = 2p$: $X_v \overset{\text{unif}}{\sim} \{\pm 1\}$ independently of X_u

under this event we'll say edge "doesn't break"

under this event we'll say edge "breaks"

- Observation: suppose that $X_V \in \{\pm 1\}^V$ is a sample from broadcast process on tree

then $X_V \sim P_\theta(x) \propto \exp\left(\sum_{(u,v)\in E} \theta \cdot x_u x_v\right)$

where $\theta = \frac{1}{2}\log\frac{1-p}{p}$

tree structured Ising model w/ same interaction strength on all edges & no external fields

$$\left[P(X_v = x_v) \right] = \overset{\text{root}}{\frac{1}{2}} \cdot \prod_{(u,v)\in E} (1-p)^{\mathbb{1}_{x_u = x_v}} \cdot p^{\mathbb{1}_{x_u \neq x_v}}$$

$$= \frac{1}{2} \prod_{(u,v)\in E} (1-p)^{\frac{1+x_u x_v}{2}} \cdot p^{\frac{1-x_u x_v}{2}}$$

$$= \frac{1}{2} \prod_{(u,v)\in E} \exp\left(\frac{1}{2}(1+x_u x_v)\log(1-p) + \frac{1}{2}(1-x_u x_v)\cdot \log p\right)$$

$$\propto \prod_{(u,v)\in E} \exp\left(\frac{1}{2} x_u x_v \log \frac{1-p}{p}\right)$$

$$\propto \exp\left(\sum_{(u,v)\in E} \underbrace{\frac{1}{2}\log\frac{1-p}{p}}_{\equiv\ \theta} x_u x_v\right)$$

Things to note: $p = \frac{1}{2} \Rightarrow \theta = 0$ all nodes independent

$p = 0 \Rightarrow \theta = +\infty$ all nodes infinitely correlated

$$X_v = \begin{cases} \vec{1}, \text{ wpr } 1/2 \\ -\vec{1}, \text{ wpr } 1/2 \end{cases}$$

$\rightsquigarrow |L(h)| = 2^h$

- Question: Suppose $L(h)$ are nodes at depth h.
  How much information about x_ρ does $X_{L(h)}$ contain?

  Observation 1: with probability $(2p)^2$, children of the root & therefore all their descendents are independent of root!

  If this happens already at depth 1, what happens as $h \to \infty$?

  To study this formally:

  suppose $\mu_h^+(\cdot) = \Pr[X_{L(h)} = \cdot \mid X_\rho = +1]$
  & $\mu_h^-(\cdot) = \Pr[X_{L(h)} = \cdot \mid X_\rho = -1]$

  Le Cam $\Rightarrow$ given $X_{L(h)}$ the best estimator of the root state has error

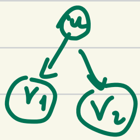
$$\frac{1}{2}\left(1 - TV(\mu_h^+, \mu_h^-)\right)$$

  Def: We'll say that the root reconstruction problem is solvable if

$$\liminf_{h \to \infty} TV(\mu_h^+, \mu_h^-) > 0$$

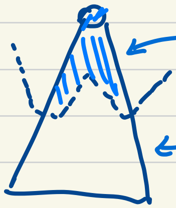
  o.w. we call it unsolvable

Observation 2: Reconstruction problem is unsolvable if $\underbrace{2\cdot\lambda}_{\text{expected \# of children who copy } x_u} < 1$



$\Leftrightarrow 1-2p < \frac{1}{2} \Leftrightarrow \frac{1}{2} < 2p \Leftrightarrow \frac{1}{4} < p$

By birth-death process analysis if $2\lambda < 1$ the state of the root will "die" w pr 1 on an infinite tree



nodes that copy x_ρ

nodes that are independent of x_ρ

Question: Is it solvable when $2\lambda > 1$ $(\Leftrightarrow p < \frac{1}{4})$?

Theorem. The root reconstruction problem is solvable iff $2\lambda^2 > 1$ $(\Leftrightarrow p < \frac{1}{4}(\underbrace{2-\sqrt{2}}_{\approx 0.5858}))$

Proof: we will only show sufficiency i.e. $2\lambda^2 > 1 \Rightarrow$ solvable

- We'll show that the majority of the leaf values maintains reconstruction error $< \frac{1}{2}$ as $h \to \infty$
- Consider:

$$Z_h \triangleq \frac{1}{2^h\cdot\lambda^h}\sum_{v\in L_h} x_v$$

$\to \hat{\mu}_h^+$: dist'n of Z_h conditioning on $\rho = +1$

$\to \hat{\mu}_h^-$: dist'n of Z_h conditioning on $\rho = -1$

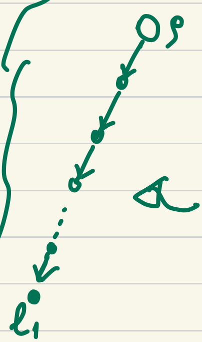
① **Lemma 1:** $\mathbb{E}[Z_h \mid X_\rho] = X_\rho$

Proof: $$\mathbb{E}\left[Z_h \mid X_\rho\right] = \frac{1}{\lambda^h}\,\mathbb{E}\left[X_{\ell_1} \mid X_\rho\right]$$

← left-most leaf at depth h

$$= \frac{1}{\lambda^h}\cdot\left(X_\rho\cdot\lambda^h + 0\cdot(1-\lambda^h)\right) = X_\rho$$

$\boxtimes$



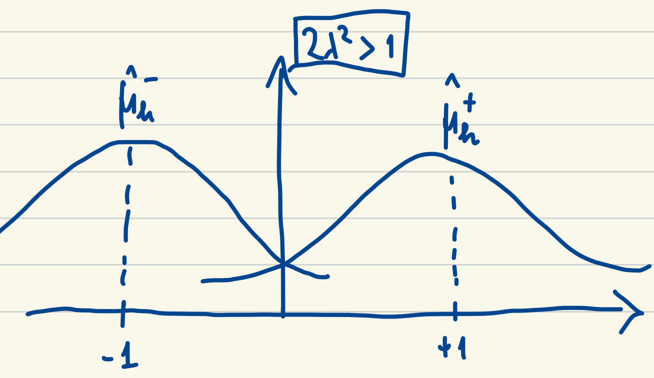
← If any of these edges "break" X_{ℓ_1} is independent of X_ρ

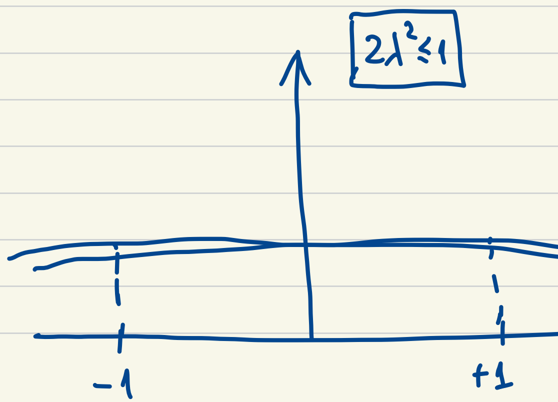
If none of the edges "break" $X_{\ell_1} = X_\rho$

② **Lemma 2:** $$\mathrm{Var}[Z_h] \xrightarrow[h\to\infty]{} \begin{cases} \dfrac{1/2}{1-(2\lambda^2)^{-1}}, & \text{if } 2\lambda^2 > 1 \\ +\infty, & \text{if } 2\lambda^2 \le 1 \end{cases}$$

Proof: postpone for a bit

mental picture:





③ Using Lemma 1 & Lemma 2 to prove sufficiency when $2\lambda^2 > 1$

recall definitions of $\mu_h^+, \mu_h^-, \hat{\mu}_h^+, \hat{\mu}_h^-$

claim: $TV(\hat{\mu}_h^+, \hat{\mu}_h^-) \leq TV(\mu_h^+, \mu_h^-)$

proof: easy $Z_h = f(X_{L_h})$

average f'n

take optimal coupling of $X_{L_h}^+ \sim \mu_h^+$ and $X_{L_h}^- \sim \mu_h^-$

implies a coupling of $Z_h^+ \sim \hat{\mu}_h^+$ and $Z_h^- \sim \hat{\mu}_h^-$

under this coupling: $\underbrace{\Pr[Z_h^+ \neq Z_h^-]}_{\geq TV(\hat{\mu}_h^+, \hat{\mu}_h^-)} \leq \Pr\underbrace{[X_{L_h}^+ \neq X_{L_h}^-]}_{= TV(\mu_h^+, \mu_h^-)}$

b.c. this is some coupling of $\hat{\mu}_h^+$ and $\hat{\mu}_h^-$

b.c. optimal coupling of μ_h^+, μ_h^-

so to show solvability, suffices to lower bound

$$TV(\hat{\mu}_h^+, \hat{\mu}_h^-) = \frac{1}{2}\sum_z |\hat{\mu}_h^+(z) - \hat{\mu}_h^-(z)|$$

$$\geq \sum_z \left(\frac{\hat{\mu}_h^+(z) - \hat{\mu}_h^-(z)}{2\hat{\mu}_h(z)}\right)^2 \hat{\mu}_h(z)$$

used $\frac{|\hat{\mu}_h^+(z) - \hat{\mu}_h^-(z)|}{2\hat{\mu}_h(z)} \leq 1$

which follows from $\hat{\mu}_h(z) = \frac{1}{2}\hat{\mu}_h^+(z) + \frac{1}{2}\hat{\mu}_h^-(z)$

Cauchy-Schwartz

$$\geq \frac{\left(\sum_z z\left(\frac{\hat{\mu}_h^+(z)-\hat{\mu}_h^-(z)}{2\hat{\mu}_h(z)}\right)\hat{\mu}_h(z)\right)^2}{\sum_z z^2\hat{\mu}_h(z)}$$

$$= \frac{1}{4}\frac{\left(\mathbb{E}_h^+(z_h)-\mathbb{E}_h^-(z_h)\right)^2}{\mathrm{Var}(z_h)}$$

$$= \frac{1}{\mathrm{Var}(z_h)}$$

So: $\liminf_{h\to\infty} TV(\mu_h^+,\mu_h^-)$

$$\geq \liminf_{h\to\infty} TV(\hat{\mu}_h^+,\hat{\mu}_h^-)$$

claim

$$\geq \liminf_{h\to\infty}\frac{1}{\mathrm{Var}(z_h)} = 2\left(1-(2\lambda^2)^{-1}\right) > 0$$

above calculations

lemma 2

b.c. $2\lambda^2>1$

④ proof of Lemma 2:

$$\mathrm{Var}[z_h] = \mathrm{Var}\left[\mathbb{E}[z_h|X_\rho]\right] + \mathbb{E}\left[\mathrm{Var}[z_h|X_\rho]\right]$$

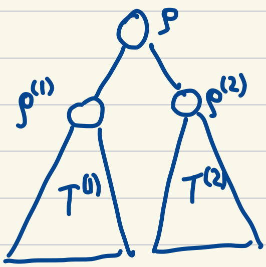
$$= \mathrm{Var}[X_\rho] + \frac{1}{2}\mathrm{Var}[z_h|X_\rho=+1] + \frac{1}{2}\mathrm{Var}[z_h|X_\rho=-1]$$

lemma 1

using $\mathrm{Var}[X_\rho]=1$
& symmetry

$$= 1 + \mathrm{Var}\left[Z_h \mid X_\rho = +1\right]$$

To understand $\mathrm{Var}\left[Z_h \mid X_\rho = +1\right]$ consider



$$Z_h = Z_h^{(1)} + Z_h^{(2)}$$

↳ contribution of leaves of T_1 to the sum

since $Z_h^{(1)}, Z_h^{(2)}$ are ind. conditioning on X_ρ

$$\mathrm{Var}\left[Z_h \mid X_\rho = +1\right] = \mathrm{Var}\left[Z_h^{(1)} \mid X_\rho = +1\right] + \mathrm{Var}\left[Z_h^{(2)} \mid X_\rho = +1\right]$$

symmetry

$$= 2\,\mathrm{Var}\left[Z_h^{(1)} \mid X_\rho = +1\right]$$

$$= 2 \cdot \left(\mathbb{E}\left[\left(Z_h^{(1)}\right)^2 \mid X_\rho = +1\right] - \left(\mathbb{E}\left[Z_h^{(1)} \mid X_\rho = +1\right]\right)^2 \right)$$

Now note: $\mathbb{E}\left[Z_h^{(1)} \mid X_\rho = +1\right] = \frac{1}{2}$ (by symmetry b.w. $Z_h^{(1)}$ & $Z_h^{(2)}$ they have the same expectation cond. on $X_\rho = +1$ & the sum of these expectations is $+1$)

Moreover:

$$\mathbb{E}\left[\left(Z_h^{(1)}\right)^2 \mid X_\rho = +1\right] = (1-p) \cdot \mathbb{E}\left[\left(Z_h^{(1)}\right)^2 \mid X_{\rho^{(1)}} = +1\right]$$
$$+ \, p \cdot \mathbb{E}\left[\left(Z_h^{(1)}\right)^2 \mid X_{\rho^{(1)}} = -1\right]$$

$$= (1-p)\cdot \frac{1}{(2\lambda)^2}\, \mathbb{E}\left[Z_{h-1}^2 \,\middle|\, X_p = +1 \right] + p \cdot \frac{1}{(2\lambda)^2} \cdot \underbrace{\mathbb{E}\left[Z_{h-1}^2 \,\middle|\, X_p = -1 \right]}_{= \, \mathbb{E}\left[Z_{h-1}^2 \,\middle|\, X_p = +1 \right]}$$

$$= \frac{1}{(2\lambda)^2} \cdot \mathbb{E}\left[Z_{h-1}^2 \,\middle|\, X_p = +1 \right]$$

putting everything together:

$$\mathrm{Var}\left[Z_h \right] = 1 + 2 \cdot \left(\frac{1}{(2\lambda)^2}\, \mathbb{E}\left[Z_{h-1}^2 \,\middle|\, X_p = +1 \right] - \frac{1}{4} \right)$$

$$= 1 + \frac{1}{2\lambda^2} \cdot \mathbb{E}\left[Z_{h-1}^2 \,\middle|\, X_p = +1 \right] - \frac{1}{2}$$

$$= \frac{1}{2} + \frac{1}{2\lambda^2} \cdot \left(\mathrm{Var}\left[Z_{h-1} \,\middle|\, X_p = +1 \right] + \left(\mathbb{E}\left[Z_{h-1} \,\middle|\, X_p = +1 \right] \right)^2 \right)$$

$$= \frac{1}{2} + \frac{1}{2\lambda^2} \cdot \underbrace{\left(\mathrm{Var}\left[Z_{h-1} \,\middle|\, X_p = +1 \right] + 1 \right)}_{= \, \mathrm{Var}\left[Z_{h-1} \right] \text{ as shown earlier}}$$

$$= \frac{1}{2} + \frac{1}{2\lambda^2} \cdot \mathrm{Var}\left[Z_{h-1} \right]$$

Solving recursion gives result $\boxtimes$

[

$$
\begin{aligned}
g(h) &= a + b \cdot g(h-1) \\
&= a + b\left(a + b\, g(h-2)\right) \\
&= a + b \cdot a + b^2 \cdot g(h-2) \\
&= a + ba + b^2 a + b^3 g(h-3) \\
&= a\left(1 + b + b^2 + \dots + b^{k-1}\right) + b^k \cdot g(h-k) \\
&\underset{k=h}{=} a\left(1 + b + \dots + b^{h-1}\right) + b^h g(0) \\
&= a \frac{b^h - 1}{b - 1} + b^h \cdot g(0)
\end{aligned}
$$

$$
\mathrm{Var}[Z_h] = \frac{1}{2} \cdot \frac{1 - \left(\frac{1}{2\lambda^2}\right)^h}{1 - \frac{1}{2\lambda^2}} + \left(\frac{1}{2\lambda^2}\right)^h \xrightarrow[h\to\infty]{} \begin{cases} \dfrac{\frac{1}{2}}{1 - \frac{1}{2\lambda^2}} & 2\lambda^2 > 1 \\ +\infty, & 2\lambda^2 < 1 \end{cases}
$$

If $2\lambda^2 = 1$

$$
g(h) = a + g(h-1) = h \cdot a + g(0)
$$

$$
\mathrm{Var}[Z_h) = h \cdot \frac{1}{2} + 1 \rightarrow +\infty
$$

]

Post Mortem:

high temperature $2\lambda^2 < 1 \iff p \geq \frac{1}{2}\left(1-\frac{1}{\sqrt{2}}\right)$

$$\Updownarrow$$

$$\theta \leq \frac{1}{2}\log\frac{\sqrt{2}+1}{\sqrt{2}-1}$$

is BAD: prevents root reconstruction

low temperature $2\lambda^2 > 1 \iff p < \frac{1}{2}\left(1-\frac{1}{\sqrt{2}}\right) \iff \theta > \frac{1}{2}\log\frac{\sqrt{2}+1}{\sqrt{2}-1}$

is GOOD: allows root reconstruction

$\limsup_{h\to\infty} \Pr[\text{incorrect reconstruction}] =$

$$\limsup_{h\to\infty} \frac{1}{2}\left(1 - TV(\mu_h^+, \mu_h^-)\right) = \frac{1}{2} - \left(1 - \frac{1}{2\lambda^2}\right) < \frac{1}{2}$$

Extensions:

- d-ary tree, binary symmetric channel: threshold at $b \cdot \lambda^2 = 1$
- more general setting: q possible states per node
  d-ary tree
  every edge: channel M

- for $q=2$: threshold at $d\,\lambda_2(M)^2 = 1$

$d \cdot \lambda_2(M)^2 > 1$ sufficient [Kesten-Stigum '66]

necessary [Bleher-Ruiz-Zagrebnov '95]

- larger q: $d \cdot \lambda_2(M)^2 > 1$ sufficient [Kesten-Stigum '66]

  but not necessary!

  [Mossel '01] [Sly '09] give examples for $q \geq 5$ for which $\lambda_2(M) = 0$ yet reconstruction is possible!

  for $q = 3, 4$: KS-bound is tight if $d > d_0(M)$ [Mossel-Sly-Sohn '23]

- for any q, if reconstruction algo only uses the <u>counts</u> of how many leaves are $+1$ or -1 then KS-bound is tight! [Mossel-Peres '03]

- <u>Conjecture</u> [Koehler-Mossel '22]: not just linear fn's but also low-degree polynomials of the X_{L_h} also fail below K-S bound!