MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.265/15.070J Lecture 2 Feb 15, SP17
Lecturer: Guy Bresler Scribe notes by Dheeraj Nagaraj

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They are posted to serve class purposes.*

Broadcasting on Trees

Content.

1. Broadcasting on Trees
2. Reconstruction Problem
3. Percolation Connection
4. Simple Estimator
5. "2nd Moment Method for Total Variation"

1 Broadcasting on Trees

Consider the infinite tree $T = \widehat{\mathbb{T}}_d = (V, E)$ with d children at each node. Let ∂n denote the vertices at distance n from the root node ρ.

We define a stochastic process over the vertices of T as follows. Assign a random variable $\sigma_v \in \{-1, 1\}, \forall v \in V$. Choose σ_ρ uniformly at random. For $\rho \neq v \in V$, if u is the parent of v, $\sigma_v = \sigma_u$ w.p $1 - \epsilon$ and $\sigma_v = -\sigma_u$ w.p. ϵ. Every parent-child path can be seen as a markov chain with a transition matrix

$$M = \begin{bmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{bmatrix}$$

We ask the following questions: Suppose we observe all the random variables at level n (denoted by $\sigma_{\partial n} \in \{-1,1\}^{\partial n}$), how well can we guess the root value?

Suppose the optimal algorithm guesses the correct answer with probability $\frac{1}{2} + \Delta$ where $\Delta = \Delta(n, T, \epsilon)$, for what values of ϵ can we expect $\lim_{n\to\infty} \Delta(n, T, \epsilon) > 0$? This problem has a lot of applications in analysis of Phylogenetic trees and extremal Gibbs measure.

We identify the trivial cases :

1. $\epsilon = 0 \Rightarrow \Delta = 1$

2. $\epsilon = \frac{1}{2} \Rightarrow \Delta = 0$

We have the following Theorem for binary trees.

Theorem 1. *Let* $2\theta_c^2 = 1$ *and* $\epsilon_c = \frac{1-\theta_c}{2}$. *Then,*

1. *if* $\epsilon > \epsilon_c$
$$\lim_n \Delta(\widehat{\mathbb{T}}_2, n, \epsilon) = 0$$

2. *if* $\epsilon < \epsilon_c$
$$\lim_n \Delta(\widehat{\mathbb{T}}_2, n, \epsilon) > 0$$

For a $d-$ary tree, the corresponding result is obtained when $d\theta_c^2 = 1$. We only prove the second statement of Theorem 1 for now; this bound on its own is known as the Kesten-Stigum bound.We need a few more concepts before proceeding with the proof.

2 Hypothesis Testing and Reconstruction

We look for a function $\widehat{\sigma}_\rho(n) : \{-1, 1\}^{\partial n} \to \{-1, 1\}$ such that given the values of $\sigma_{\partial n}$, it guesses the value of σ_ρ.

1. If $\sigma_\rho = 1 : \sigma_{\partial n}$ has a distribution given by the function $\mathbb{P}^+_{\partial n} := \mathbb{P}(\sigma_{\partial n} \in A | \sigma_\rho = 1)$

2. If $\sigma_\rho = -1 : \sigma_{\partial n}$ has a distribution given by the function $\mathbb{P}^-_{\partial n} := \mathbb{P}(\sigma_{\partial n} \in A | \sigma_\rho = -1)$

We treat this problem in a more general way. Suppose we have random variables $X \in \mathcal{X}$ (called 'observation') and $H \in \{0, 1\}$ (called 'Hypothesis'). (Assume $\mathcal{X}$ is a discrete set)H is distributed uniformly at random and

1. over $\{H = 0\}$, $X \stackrel{d}{=} p$

2. over $\{H = 1\}$, $X \stackrel{d}{=} q$

Therefore, by Bayes' theorem, $X \stackrel{d}{=} \frac{p+q}{2}$ We define a test $\Psi : \mathcal{X} \to \{0, 1\}$ which is a function which guesses the value of H given the value of $X \in \mathcal{X}$. We look for Ψ such that the probability of error in prediction is minimized.

2.1 A Lower Bound for $\mathbb{P}[\text{error}]$

Define $A = \{x : \Psi(x) = 0\}$. Clearly,

$$\mathbb{P}[\text{error}] = \frac{1}{2}\mathbb{P}[X \in A^c | H = 0] + \frac{1}{2}\mathbb{P}[X \in A | H = 1] \tag{1}$$

$$= \frac{1}{2}\left(p(A^c) + q(A)\right) \tag{2}$$

$$= \frac{1}{2} - \frac{1}{2}\left(p(A) - q(A)\right) \tag{3}$$

$$\geq \frac{1}{2} - \sup_A |p(A) - q(A)| = \frac{1}{2} - \frac{1}{2}\mathrm{d_{TV}}(p.q) \tag{4}$$

Where $\mathrm{d_{TV}}(p.q) := \sup_A |p(A) - q(A)|$, is called the total variation distance.

Claim 1 (Exercise)**.** $\mathrm{d_{TV}}$ *is a metric on the space of distribution over* $\mathcal{X}$. *Further, if* $\mathcal{X}$ *is discrete, then* $A^* := \{x \in \mathcal{X} : p(x) > q(x)\} \subseteq \mathcal{X}$ *satisfies* $d_{TV}(p, q) = p(A^*) - q(A^*)$.

Clearly, from the lower bound of the probability of error, we conclude that the function $\Psi^* = \mathbb{1}_{A^*}$ minimizes the probability of error and hence is the optimum predictor. Therefore,

$$\mathbb{P}[\text{error } \Psi^*] = \frac{1}{2} - \frac{1}{2}\mathrm{d_{TV}}(p, q)$$

Claim 2 (Exercise)**.**

$$\mathrm{d_{TV}}(p, q) = \frac{1}{2}\sum_{x \in \mathcal{X}} |p(x) - q(x)|$$

2.2 Optimal Reconstruction

Define

$$\widehat{\sigma}_\rho^{\mathrm{ML}}(s) = \begin{cases} +1 \text{ if } \mathbb{P}^+_{\partial n}(s) > P^-_{\partial n}(s) \\ -1 \text{ otherwise} \end{cases}$$

By the preceding discussion, the ML estimator (Maximum Likelihood) should give us the minimum possible probability of error. Since this estimator is hard to analyse, we use a sub-optimal estimator

$$\widehat{\sigma}^{\mathrm{MAJ}} := \mathrm{sign}(S_n),$$

where $S_n = \sum_{v \in \partial n} \sigma_v$.

Claim 3 (Exercise)**.**

$$\widehat{\sigma}^{\mathrm{ML}} \neq \widehat{\sigma}^{\mathrm{MAJ}}$$

Clearly, $\Delta^{\mathrm{MAJ}} = \mathrm{d}_{\mathrm{TV}}(p^{+}_{S_n}, p^{-}_{S_n})$. To prove the second part of Theorem 1 with S_n, it is sufficient to show that $\lim_n \Delta^{\mathrm{MAJ}} > 0$.

3 Random Cluster Representation and Connection to Percolation

We couple the stochastic process over the signs of the vertices of the tree with the percolation one on the edges of the tree (i.e, define a joint process) such that the marginals of the process over the edges gives the broadcasting process defined above. We make the following observation:

$$M = \begin{bmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{bmatrix} = \theta I + (1-\theta) \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Where $\theta = 1 - 2\epsilon$. The identity matrix corresponds to copying the same vertex state as the parent and the second matrix correponds to picking the child state uniformly and independent of the parent. Therefore we define the following percolation model:

1. Do percolation on the tree with probability θ of any edge being open.
2. Choose $\sigma_\rho \in \{-1, 1\}$ uniformly at random.
3. If there is an open edge between a parent u and a child v, then $\sigma_v = \sigma_u$ -i.e, copy the value of the parent.
4. If the edge between u and v is closed, then pick σ_v independently and uniformly at random.

Therefore, the process here is over the product of two spaces - $\{-1,1\}^V \times \{0,1\}^E$. It is easy to show that, since all edges are independently open or closed, the marginal distribution over $\{-1,1\}^V$ is the same as in the broadcasting model.

4 Second Moment Method Bound

Recall that $\widehat{\sigma}^{\mathrm{MAJ}}_\rho = \mathrm{sgn}(S_n)$.

Proposition 2.

$$\mathbb{E}^{\pm}[S_n] = \pm(2\theta)^n$$

Where $\mathbb{E}^{\pm}$ is the expectation given $\sigma_\rho = \pm 1$

Proof. Let $v \in \partial n$.

$$\mathbb{E}^+(\sigma_v) = \mathbb{E}^+[\sigma_v | v \leftrightarrow \rho]\mathbb{P}[v \leftrightarrow \rho] + \mathbb{E}^+[\sigma_v | v \not\leftrightarrow \rho]\mathbb{P}[v \not\leftrightarrow \rho]$$

Clearly, $\mathbb{E}^+[\sigma_v | v \not\leftrightarrow \rho] = 0$ and $\mathbb{E}^+[\sigma_v | v \leftrightarrow \rho] = 1$ and $\mathbb{P}[v \leftrightarrow \rho] = \theta^n$. Therefore $\mathbb{E}^+(\sigma_v) = \theta^n$. Since there are 2^n vertices in ∂n, we conclude that $\mathbb{E}^+(S_n) = (2\theta)^n$. By similar arguments, can conclude $\mathbb{E}^-$ result.

□

Proposition 3.

$$\frac{\mathrm{var}(S_n)}{(\mathbb{E}^+[S_n])^2} \to \begin{cases} \frac{1}{2}\frac{1}{1-(2\theta^2)^{-1}} \text{ if } 2\theta^2 > 1 \\ 0 \text{ otherwise} \end{cases}$$

Proof.

$$\begin{aligned} \mathbb{E}^+[S_n^2] &= \sum_{u,v \in \partial n} \mathbb{E}(\sigma_u \sigma_v) \\ &= 2^n + \sum_{u \neq v} \mathbb{E}(\sigma_u \sigma_v) \\ &\overset{\text{Ex}}{=} 2^n + \frac{1}{2} 2^{2n} \theta^{2n} \sum_{m=0}^{n-1} (2\theta^2)^{-m} \end{aligned}$$

We get the last relation by considering pairs with nearest common ancestor at level m, probability that this path is connected to both the considered vertices and then summing up over m. Now by symmetry,

$$\begin{aligned} \mathrm{var}(S_n) &= \mathbb{E}(S_n^2) - (\mathbb{E}(S_n))^2 \\ &= \mathbb{E}(S_n^+)^2 \end{aligned}$$

We use Proposition 2 to conclude the result. □

We use Chebyshev's inequality to (unsuccessfully) prove the result. Suppose that $2\theta^2 > 1$. Then,

$$\begin{aligned}
\mathbb{P}[\text{error}] &= \frac{1}{2}\left(\mathbb{P}^+[S_n \leq 0] + \mathbb{P}^-[S_n > 0]\right) \\
&= \mathbb{P}^+[S_n \leq 0] \\
&\leq \mathbb{P}^+[|S_n - \mathbb{E}^+(S_n)| \geq \mathbb{E}^+(S_n)] \\
&\leq \frac{\text{var}^+(S_n)}{(\mathbb{E}(S_n))^2} \\
&\to \frac{1}{2}\frac{1}{1-(2\theta^2)^{-1}} - 1
\end{aligned}$$

RHS is $< \frac{1}{2}$ iff $\theta > \sqrt{\frac{3}{4}}$. This doesn't give us the bound we claimed in Theorem 1.

5 Second Moment Bound for Total Variation

Claim 4. *If* $2\theta^2 > 1$ *then,*

$$\lim_n d_{\text{TV}}(P^+_{S_n}, P^-_{S_n}) \geq 2(1-(2\theta^2)^{-1})$$

We prove a slightly more general result. Consider the hypothesis testing problem of Section 2. We have the following inequality.

Proposition 4.

$$d_{\text{TV}}(p,q) \geq \frac{1}{4}\frac{(\mathbb{E}_p(X) - \mathbb{E}_q(X))^2}{\text{var}(X)}$$

Proof. Let $\tilde{p} := \frac{1}{2}(p+q)$ give the distribution of X. Over the set $\{z : \tilde{p}(z) = 0\}$, $p(z) = 0$ and $q(z) = 0$. Therefore, in summations below, we exclude these values of z from $\mathcal{X}$

$$\begin{aligned}
d_{\text{TV}}(p,q) &= \frac{1}{2}\sum_z |p(z) - q(z)| \\
&= \sum_z \frac{|p(z)-q(z)|}{2\tilde{p}(z)}\tilde{p}(z)
\end{aligned}$$

Define

$$f(z) = \begin{cases} \frac{|p(z)-q(z)|}{2\tilde{p}(z)} \text{ if } \tilde{p}(z) > 0 \\ 0 \text{ otherwise} \end{cases}$$

Define the random variable $Y = f(X)$. It is clear that $0 \leq f(z) \leq 1$. Therefore, $\mathbb{E}(Y^2) = \sum_z f(z)^2 \tilde{p}(z) \leq \sum_z f(z)\tilde{p}(z) = d_{\text{TV}}(p,q)$ By Cauchy Schwarz inequality,

$$
\begin{aligned}
\mathbb{E}(Y^2) &\geq \frac{(\mathbb{E}(XY))^2}{\mathbb{E}(X^2)} \\
&= \frac{(\sum_z z\tilde{p}(z)\frac{|p(z)-q(z)|}{2\tilde{p}(z)})^2}{\sum_z z^2\tilde{p}(z)} \\
&\geq \frac{(\sum_z z\tilde{p}(z)\frac{p(z)-q(z)}{2\tilde{p}(z)})^2}{\sum_z z^2\tilde{p}(z)} \\
&= \frac{1}{4}\frac{(\mathbb{E}_p(X) - \mathbb{E}_q(X))^2}{\text{var}(X)}
\end{aligned}
$$

And we conclude the said result. $\square$

We immediately conclude that $\lim_{n\to\infty} d_{\text{TV}}(P^+, P^-) \geq 2(1 - (2\theta^2)^{-1})$ when $2\theta^2 > 1$. Choose A_1 to be the set where total variation between S_n^+ and S_n^- is achieved. Let the estimator be $\Psi_1(n) = \mathbb{1}_{A_1}$. Clearly,

$$
\begin{aligned}
\mathbb{P}_{\Psi_1}[\text{error}] &= \frac{1}{2} - \frac{1}{2}d_{\text{TV}}(P^+, P^-) \\
\Rightarrow \lim_{n\to\infty} \mathbb{P}_{\Psi_1}[\text{error}] &< \frac{1}{2}
\end{aligned}
$$

proving our result.