Lecture 13 : Kesten-Stigum bound

MATH285K - Spring 2010 *Lecturer: Sebastien Roch*

References: [EKPS00, Mos01, MP03, BCMR06].

Previous class

DEF 13.1 (Ancestral reconstruction solvability) *Let μ_h^+ be the distribution μ_h conditioned on the root state σ_0 being $+1$, and similarly for μ_h^-. We say that the ancestral reconstruction problem (under the CFN model) for $0 < p < 1/2$ is solvable if*

$$\liminf_h \|\mu_h^+ - \mu_h^-\|_1 > 0,$$

otherwise the problem is unsolvable.

THM 13.2 (Solvability) *Let $\theta_* = 1 - 2p_* = 1/\sqrt{2}$. Then when $p < p_*$ the ancestral reconstruction problem is solvable.*

1 Kesten-Stigum bound

The previous theorem was proved by showing that majority is a good root estimator up to $p = p_*$. Here we show that this result is best possible. Of course, majority is not the best root estimator: in general its error probability can be higher than maximum likelihood. (See Figure 3 in [EKPS00] for an insightful example where majority and maximum likelihood differ.) However, it turns out that the critical threshold for majority, called the *Kesten-Stigum bound*, coincides with the critical threshold of maximum likelihood—*in the CFN model.* Note that the latter is not true for more general models [Mos01].

THM 13.3 (Tightness of Kesten-Stigum Bound) *Let $\theta_* = 1 - 2p_* = 1/\sqrt{2}$. Then when $p \geq p_*$ the ancestral reconstrution problem is not solvable.*

Along each path from the root, information is lost through mutation at exponential rate—maeasured by $\theta = 1 - 2p$. Meanwhile, the tree is growing exponentially and information is duplicated—measured by the branching ratio $b = 2$. These two forces balance each other out when $b\theta^2 = 1$, the critical threshold in the theorem.

To prove Theorem 13.3 we analyze the maximum likelihood estimator. Let $\mu_h(s_0|\mathbf{s}_h)$ be the conditional probability of the root state s_0 given the states $\mathbf{s}_h$ at level h. It will be more convenient to work with the following related quantity

$$Z_h = \mu_h(+|\boldsymbol{\sigma}_h) - \mu_h(-|\boldsymbol{\sigma}_h) = \frac{1}{2\mu_h(\boldsymbol{\sigma}_h)}[\mu_h^+(\boldsymbol{\sigma}_h) - \mu_h^-(\boldsymbol{\sigma}_h)] = 2\mu_h(+|\boldsymbol{\sigma}_h) - 1,$$

which, as a function of $\boldsymbol{\sigma}_h$, is a random variable. Note that $\mathbb{E}[Z_h] = 0$ by symmetry. It is enough to prove a bound on the variance of Z_h.

LEM 13.4 *It holds that*

$$\|\mu_h^+ - \mu_h^-\|_1 \leq 2\sqrt{\mathbb{E}[Z_h^2]}.$$

Proof: By Bayes' rule and Cauchy-Schwarz

$$\begin{aligned}\sum_{\mathbf{s}_h} |\mu_h^+(\mathbf{s}_h) - \mu_h^-(\mathbf{s}_h)| &= \sum_{\mathbf{s}_h} 2\mu_h(\mathbf{s}_h)\,|\mu_h(+|\mathbf{s}_h) - \mu_h(-|\mathbf{s}_h)| \\ &= 2\mathbb{E}|Z_h| \\ &\leq 2\sqrt{\mathbb{E}[Z_h^2]}.\end{aligned}$$

■

Let $\bar{z}_h = \mathbb{E}[Z_h^2]$. The proof of Theorem 13.3 will follow from

$$\lim_h \bar{z}_h = 0.$$

We apply the same type of recursive argument we used for the analysis of majority: we condition on the root to exploit conditional independence; we apply the Markov channel on the top edge.

2 Distributional recursion

We first derive a recursion for Z_h. Let $\dot{\boldsymbol{\sigma}}_h$ be the states at level h below the first child of the root and let $\dot{\mu}_h$ be the distribution of $\dot{\boldsymbol{\sigma}}_h$. Define

$$\dot{Z}_h = \dot{\mu}_h(+|\dot{\boldsymbol{\sigma}}_h) - \dot{\mu}_h(-|\dot{\boldsymbol{\sigma}}_h),$$

where $\dot{\mu}_h(s_0|\dot{\mathbf{s}}_h)$ is the conditional probability that the root is s_0 given that $\dot{\boldsymbol{\sigma}}_h = \dot{\mathbf{s}}_h$. Similarly, denote with a double dot the same quantities with respect to the subtree below the second child of the root.

LEM 13.5 *It holds pointwise that*

$$Z_h = \frac{\dot{Z}_h + \ddot{Z}_h}{1 + \dot{Z}_h \ddot{Z}_h}.$$

Proof: Using $\mu_h^+(\mathbf{s}_h) = \dot{\mu}_h^+(\dot{\mathbf{s}}_h)\ddot{\mu}_h^+(\ddot{\mathbf{s}}_h)$, note that

$$\begin{aligned}
Z_h &= \frac{1}{2}\sum_{\gamma=+,-}\gamma\frac{\mu_h^\gamma(\boldsymbol{\sigma}_h)}{\mu_h(\boldsymbol{\sigma}_h)} \\
&= \frac{1}{2}\frac{\dot{\mu}_h(\dot{\boldsymbol{\sigma}}_h)\ddot{\mu}_h(\ddot{\boldsymbol{\sigma}}_h)}{\mu_h(\boldsymbol{\sigma}_h)}\sum_{\gamma=+,-}\frac{\dot{\mu}_h^\gamma(\dot{\boldsymbol{\sigma}}_h)\ddot{\mu}_h^\gamma(\ddot{\boldsymbol{\sigma}}_h)}{\dot{\mu}_h(\dot{\boldsymbol{\sigma}}_h)\ddot{\mu}_h(\ddot{\boldsymbol{\sigma}}_h)} \\
&= \frac{1}{2}\frac{\dot{\mu}_h(\dot{\boldsymbol{\sigma}}_h)\ddot{\mu}_h(\ddot{\boldsymbol{\sigma}}_h)}{\mu_h(\boldsymbol{\sigma}_h)}\sum_{\gamma=+,-}\gamma\left(\frac{1+\gamma\dot{Z}_h}{2}\right)\left(\frac{1+\gamma\ddot{Z}_h}{2}\right) \\
&= \frac{1}{4}\frac{\dot{\mu}_h(\dot{\boldsymbol{\sigma}}_h)\ddot{\mu}_h(\ddot{\boldsymbol{\sigma}}_h)}{\mu_h(\boldsymbol{\sigma}_h)}(\dot{Z}_h + \ddot{Z}_h),
\end{aligned}$$

where

$$\begin{aligned}
\frac{\mu_h(\boldsymbol{\sigma}_h)}{\dot{\mu}_h^\gamma(\dot{\boldsymbol{\sigma}}_h)\ddot{\mu}_h^\gamma(\ddot{\boldsymbol{\sigma}}_h)} &= \sum_{\gamma=+,-}\frac{1}{2}\frac{\mu_h^\gamma(\boldsymbol{\sigma}_h)}{\dot{\mu}_h^\gamma(\dot{\boldsymbol{\sigma}}_h)\ddot{\mu}_h^\gamma(\ddot{\boldsymbol{\sigma}}_h)} \\
&= \frac{1}{2}\sum_{\gamma=+,-}\left(\frac{1+\gamma\dot{Z}_h}{2}\right)\left(\frac{1+\gamma\ddot{Z}_h}{2}\right) \\
&= \frac{1}{4}(1+\dot{Z}_h\ddot{Z}_h).
\end{aligned}$$

■

Define

$$\dot{Z}_{h-1} = \dot{\mu}_{h-1}(+|\dot{\boldsymbol{\sigma}}_h) - \dot{\mu}_{h-1}(-|\dot{\boldsymbol{\sigma}}_h),$$

where $\dot{\mu}_{h-1}(s_0|\dot{\boldsymbol{\sigma}}_h)$ is the condition probability that the first child of the root is s_0 given that the states at level h below the first child are $\dot{\boldsymbol{\sigma}}_h$. Similarly,

LEM 13.6 *It holds pointwise that*

$$\dot{Z}_h = \theta\dot{Z}_{h-1}.$$

Proof: The proof is similar to that of the previous lemma and is left as an exercise.
■

3 Moment recursion

We now take expectations in the previous recursion for Z_h. Note that we need to compute the second moment. However, an important simplification arises from the following observation:

$$\begin{aligned}
\mathbb{E}^+_h[Z_h] &= \sum_{\mathbf{s}_h} \mu^+_h(\mathbf{s}_h) Z_h(\mathbf{s}_h)\\
&= \sum_{\mathbf{s}_h} \mu_h(\mathbf{s}_h)\frac{\mu^+_h(\mathbf{s}_h)}{\mu_h(\mathbf{s}_h)} Z_h(\mathbf{s}_h)\\
&= \sum_{\mathbf{s}_h} \mu_h(\mathbf{s}_h)(1+Z_h(\mathbf{s}_h))Z_h(\mathbf{s}_h)\\
&= \mathbb{E}[(1+Z_h)Z_h]\\
&= \mathbb{E}[Z_h^2],
\end{aligned}$$

so it suffices to compute the (conditioned) first moment.
Proof:(of Theorem 13.3) Using the expansion

$$\frac{1}{1+r} = 1 - r + \frac{r^2}{1+r},$$

we have that

$$\begin{aligned}
Z_h &= \theta(\dot{Z}_{h-1}+\ddot{Z}_{h-1}) - \theta^3(\dot{Z}_{h-1}+\ddot{Z}_{h-1})\dot{Z}_{h-1}\ddot{Z}_{h-1} + \theta^4\dot{Z}^2_{h-1}\ddot{Z}^2_{h-1}Z_h\\
&\leq \theta(\dot{Z}_{h-1}+\ddot{Z}_{h-1}) - \theta^3(\dot{Z}_{h-1}+\ddot{Z}_{h-1})\dot{Z}_{h-1}\ddot{Z}_{h-1} + \theta^4\dot{Z}^2_{h-1}\ddot{Z}^2_{h-1}, \quad (1)
\end{aligned}$$

where we used $|Z_h| \leq 1$. To take expectations, we need the following lemma.

LEM 13.7 *We have*

$$\mathbb{E}^+_h[\dot{Z}_{h-1}] = \theta\mathbb{E}^+_{h-1}[\dot{Z}_{h-1}],$$

and

$$\mathbb{E}^+_h[\dot{Z}^2_{h-1}] = (1-\theta)\mathbb{E}[\dot{Z}^2_{h-1}] + \theta\mathbb{E}^+_{h-1}[\dot{Z}^2_{h-1}] = \mathbb{E}[\dot{Z}^2_{h-1}] = \mathbb{E}^+_{h-1}[\dot{Z}_{h-1}].$$

Proof: For the first equality, note that by symmetry

$$\begin{aligned}
\mathbb{E}^+_h[\dot{Z}_{h-1}] &= (1-p)\mathbb{E}^+_{h-1}[\dot{Z}_{h-1}] + p\mathbb{E}^-_{h-1}[\dot{Z}_{h-1}]\\
&= (1-2p)\mathbb{E}^+_{h-1}[\dot{Z}_{h-1}].
\end{aligned}$$

The second equality is proved similarly and is left as an exercise. ■

Taking expectations in (1), using conditional independence and symmetry

$$\begin{aligned}
\bar{z}_h &\leq 2\theta^2\bar{z}_{h-1} - 2\theta^4\bar{z}^2_{h-1} + \theta^4\bar{z}^2_{h-1}\\
&= 2\theta^2\bar{z}_{h-1} - \theta^4\bar{z}^2_{h-1}.
\end{aligned}$$

■

References

[BCMR06] Christian Borgs, Jennifer T. Chayes, Elchanan Mossel, and Sébastien Roch. The Kesten-Stigum reconstruction bound is tight for roughly symmetric binary channels. In *FOCS*, pages 518–530, 2006.

[EKPS00] W. S. Evans, C. Kenyon, Y. Peres, and L. J. Schulman. Broadcasting on trees and the Ising model. *Ann. Appl. Probab.*, 10(2):410–433, 2000.

[Mos01] E. Mossel. Reconstruction on trees: beating the second eigenvalue. *Ann. Appl. Probab.*, 11(1):285–300, 2001.

[MP03] E. Mossel and Y. Peres. Information flow on trees. *Ann. Appl. Probab.*, 13(3):817–844, 2003.