CoT Information: Improved Sample Complexity under Chain-of-Thought Supervision

Awni Altabaa   Omar Montasser   John Lafferty

Statistics & Data Science, Yale University

Scan for project webpage

Motivation & High-level Goals

- **Standard supervised learning:** observe only input-output pairs; when learning very complicated functions (e.g., multi-step reasoning) this provides a weak supervision signal, requiring an intractable number of samples to learn.
- **Chain-of-Thought (CoT) supervision:** provides learner with *additional information* through observing intermediate step-by-step reasoning traces
- **Empirical Observation:** LLMs trained with CoT supervision exhibit dramatically higher accuracy on reasoning tasks
- **This work:** develops a theoretical statistical learning framework explaining *when* and *why* CoT supervision yields more rapid learning.

Theoretical Framework

Background: Standard Supervised Learning. Want to learn $h_\star : \mathcal{X} \to \mathcal{Y}$ in some function class $\mathcal{H} \subset \mathcal{Y}^{\mathcal{X}}$. Learner observes labeled examples (x_i, y_i). Learning algorithm: $\mathcal{A} : S = \{(x_i, y_i)\}_{i=1}^m \mapsto \hat{h}$.

Goal: achieve small prediction error $\mathcal{R}(\hat{h}) := \mathbb{P}\left[\hat{h}(x) \neq y\right] \leq \varepsilon$.

CoT-Supervised Learning. Each hypothesis h emits two observable outputs:

$$y = h^{\mathrm{e2e}}(x) \text{ and } z = h^{\mathrm{CoT}}(x)$$

Additional information is encoded in the CoT $z = h^{\mathrm{CoT}}(x)$, e.g., step-by-step computational trace.

Definition (CoT hypothesis class). A family of functions $\mathcal{H} \subset (\mathcal{Y} \times \mathcal{Z})^{\mathcal{X}}$, where $\mathcal{Y}$ is the output space and $\mathcal{Z}$ is the CoT space.

Example: $\mathcal{H}$ is a sequence model class (e.g., Transformers) that outputs a CoT $\boldsymbol{z} = (z_1, \ldots, z_T)$ followed by the final answer y.

CoT Learning Algorithm: A mapping $\mathcal{A} : (\mathcal{X} \times \mathcal{Y} \times \mathcal{Z})^* \to \mathcal{Y}^{\mathcal{X}}$.

Types of Error: End-to-End risk and CoT risk,

$$\mathcal{R}_{\mathcal{D}}^{\mathrm{e2e}}(h) := \mathbb{P}\left[h^{\mathrm{e2e}}(x) \neq y\right], \; \mathcal{R}_{\mathcal{D}}^{\mathrm{CoT}}(h) := \mathbb{P}\left[h(x) \neq (y, z)\right]$$

Goal: achieve small *end-to-end* prediction error $\mathcal{R}_{\mathcal{D}}^{\mathrm{e2e}}(\hat{h})$.

CoT learner observes (x, y, z), input, output, and CoT.

Evaluation metric is *end-to-end error*.

CoT Information & Improved Sample Complexity

E2E Learning is Slow for Complex Tasks. Standard end-to-end learning has a sample complexity that scales as

$$\mathcal{O}\left(\frac{\texttt{Complexity}(\mathcal{H})}{\varepsilon}\right)$$

When the function to be learned is complex, this can be *very slow*. Intuitively: *observing input-output examples alone reveals little information about the target function.*

How to measure the amount of information encoded in the CoT for distinguishing hypotheses in the hypothesis class?

Definition: CoT Information

For a CoT hypothesis class $\mathcal{H} \subset (\mathcal{Y} \times \mathcal{Z})^{\mathcal{X}}$ and distribution $\mathcal{D}$ over $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ realizable by $\mathcal{H}$, the CoT information is the function

$$\mathcal{I}_{\mathcal{D}}^{\mathrm{CoT}}(\varepsilon; \mathcal{H}) := \inf_{h \in \Delta_{\mathcal{D}}^{\mathrm{e2e}}(\varepsilon; \mathcal{H})} -\log \mathbb{P}_{x,y,z \sim \mathcal{D}}\left[(h^{\mathrm{CoT}}(x), h^{\mathrm{e2e}}(x)) = (y, z)\right],$$

where

$$\Delta_{\mathcal{D}}^{\mathrm{e2e}}(\varepsilon; \mathcal{H}) := \left\{h \in \mathcal{H} : \mathbb{P}_{x,y}\left[h^{\mathrm{e2e}}(x) \neq y\right] > \varepsilon\right\}$$

- Large $\mathcal{I}_{\mathcal{D}}^{\mathrm{CoT}}(\varepsilon) \implies$ each CoT sample carries more information $\to$ fewer samples needed.
- $\mathcal{I}_{\mathcal{D}}^{\mathrm{CoT}}(\varepsilon)/\varepsilon$ ratio interpreted as relative value of CoT sample compared to E2E sample; always ≥ 1.

The CoT Information captures the statistical rate of CoT learning

Result: Statistical Complexity of CoT Learning

CoT supervision improves the sample complexity of learning by replacing the ε-dependence with the CoT information

$$\underbrace{\Theta\left(\frac{\texttt{Complexity}(\mathcal{H})}{\varepsilon}\right)}_{\text{End-to-End Sample Complexity}} \longrightarrow \underbrace{\Theta\left(\frac{\texttt{Complexity}(\mathcal{H})}{\mathcal{I}_{\mathcal{D}}^{\mathrm{CoT}}(\varepsilon)}\right)}_{\text{CoT Sample Complexity}}$$

Result establishes complementary *upper bounds* and *lower bounds*

Informativeness of Supervision Signal	E2E Noisy	E2E Low-Noise	E2E Noiseless	CoT Supervision
Sample complexity	$\frac{1}{\varepsilon^2}$	$\frac{1}{\varepsilon^\gamma}, \; \gamma \in (1,2)$	$\frac{1}{\varepsilon}$	$\frac{1}{\mathcal{I}_{\mathcal{D}}^{\mathrm{CoT}}(\varepsilon)}$

More Information, Faster Learning $\longrightarrow$

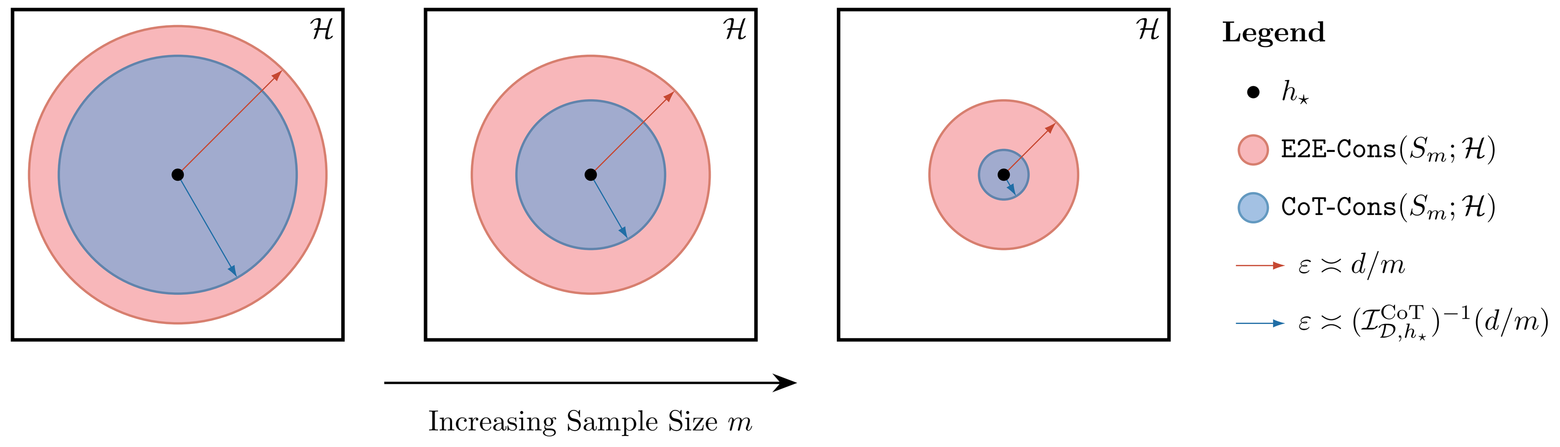


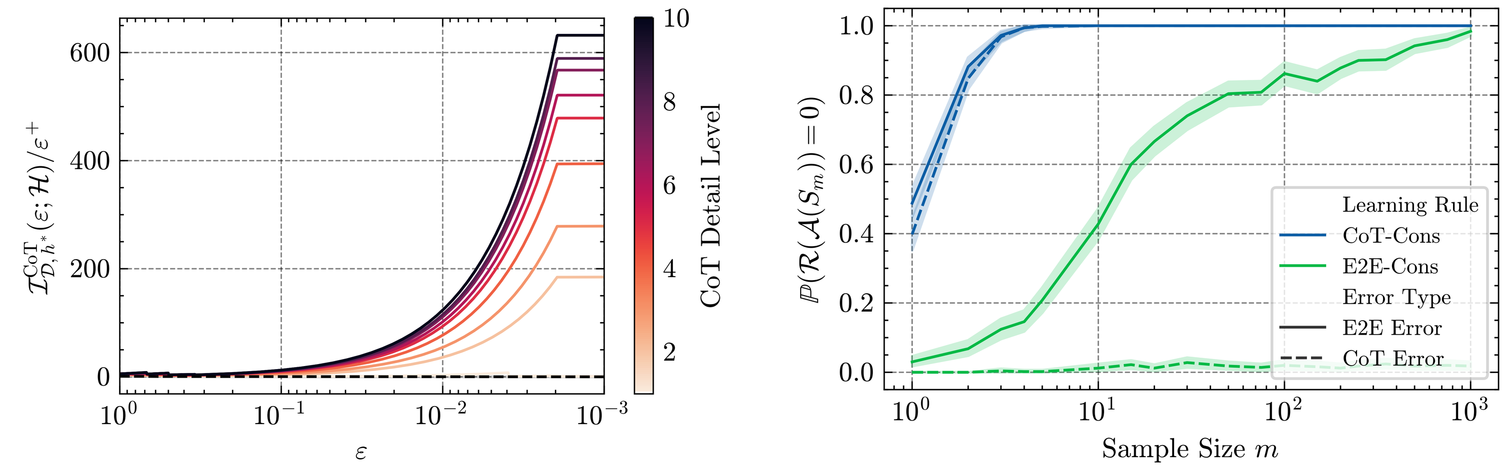
Illustration of the geometry of learning with CoT supervision compared to End-to-End supervision: the volume of the CoT constrained ball around $h_\star$ shrinks faster than the volume of the E2E constrained ball around $h_\star$

Empirical Validation of Theoretical Predictions

Learning regular language $\mathcal{L}$. End-to-end function is membership in $\mathcal{L}$. CoT is sequence of states for corresponding DFA

$$h^{\mathrm{e2e}}(x) = \mathbf{1}\{x \in \mathcal{L}\}, \; h^{\mathrm{CoT}}(x) = (z_1, \ldots, z_n),$$

where DFA has state transitions $z_t \xrightarrow{x_t} z_{t+1}$.



(a) Theoretical CoT information.   (b) Empirical learning curve.

Simulation results for learning a regular language. CoT information $\lim_{\varepsilon \to 0} \mathcal{I}_{\mathcal{D}}^{\mathrm{CoT}}(\varepsilon; \mathcal{H})/\varepsilon^+ \approx 600$ accurately predicts the empirical sample-efficiency improvement ($10^2 - 10^3\times$)

Conclusion

- Introduced the *CoT information* as a new statistical measure of supervision strength, characterizing CoT learning.
- Established statistical *upper bounds* & *lower bounds*.
- $\mathcal{I}_{\mathcal{D}}^{\mathrm{CoT}}(\varepsilon)$ provides a *quantitative metric* for the value of observing reasoning traces — practical implication: CoT annotation design.
- Offers new perspective for other problems: OOD generalization, reinforcement learning for reasoning, and scaling laws for CoT.

Contact: awni.altabaa@yale.edu