How fast can you find a good hypothesis?

Anders Aamand AA@DI.KU.DK
BARC, University of Copenhagen

Maryam Aliakbarpour MARYAMA@RICE.EDU
Rice University

Justin Y. Chen JUSTC@MIT.EDU
MIT

Sandeep Silwal SILWAL@CS.WISC.EDU
University of Wisconsin-Madison

Editors: Steve Hanneke and Tor Lattimore

Abstract

Hypothesis selection asks for a distribution close to an unknown P, given samples from P and access to n candidate hypotheses.We study the computational complexity of achieving statistically optimal sample complexity and approximation factors. [1]

Keywords: Hypothesis selection, density estimation, distribution learning

Given sample access to an unknown distribution P and sample/query access to hypotheses $\mathcal{H} = \{H_1, \dots, H_n\}$ over a (possibly infinite) domain $\mathcal{X}$, the goal is to output Q with $d_{\mathrm{TV}}(P, Q) \le C \cdot \mathsf{OPT} + \varepsilon$, where $\mathsf{OPT} = \min_i d_{\mathrm{TV}}(P, H_i)$. With $O(\log(n/\delta)/\varepsilon^2)$ samples from P, proper algorithms achieve the tight factor $C = 3$, while statistically optimal improper algorithms can achieve $C = 2$ but previously known implementations have runtimes which depend polynomially on $|\mathcal{X}|$ (Bousquet et al., 2019, 2021). We investigate whether optimal statistical guarantees admit domain-independent, near-linear time algorithms.

Mixture outputs. A natural finite-time improper strategy outputs a mixture of the hypotheses. We prove the exact optimal factor $3 - 2/n$ for such outputs: we give a lower bound for every domain-independent-sample algorithm and a matching algorithm which runs in $\mathrm{poly}(n)$ time. Thus mixtures recover the known $C = 2$ guarantee for $n = 2$ (Mahalanabis and Stefankovic, 2008), but cannot yield an absolute constant below 3 for general n.

Proper near-linear time. For proper hypothesis selection with optimal factor $C = 3$, the previous fastest algorithms run in time $O(n^2 \log(n/\delta)/\varepsilon^2)$ and $\tilde{O}(n/(\delta^3\varepsilon^3))$ (Mahalanabis and Stefankovic, 2008; Aliakbarpour et al., 2024). We improve the latter result to $\tilde{O}(n/(\delta\varepsilon^2))$ while retaining optimal sample complexity. Our algorithm reduces the task to a graph problem and solves it using ideas from sublinear graph algorithms to avoid comparing all $\Theta(n^2)$ pairs of hypotheses.

Additional resources. Given an upper bound $R \ge \mathsf{OPT}$, we obtain runtime $O(n \log^3(n/\delta)/\varepsilon^2)$, and output H_i with $d_{\mathrm{TV}}(P, H_i) \le 2\mathsf{OPT} + R + \varepsilon$. When $R = \mathsf{OPT}$, this gives $C = 3$ without a polynomial dependence on $1/\delta$. For finite domains when polynomial preprocessing of $\mathcal{H}$ is allowed, we give a data structure with $C = 3$ and $\tilde{O}(n^{2-\Omega(\varepsilon)}\varepsilon^{-3})$ query time which succeeds with high probability in n. In both settings, our algorithms utilize the extra resources to break the quadratic barrier for high probability selection with the optimal approximation.

1. Extended abstract. Full version appears at https://arxiv.org/abs/2509.03734.

© 2026 A. Aamand, M. Aliakbarpour, J.Y. Chen & S. Silwal.

References

Maryam Aliakbarpour, Mark Bun, and Adam Smith. Optimal hypothesis selection in (almost) linear time. *Advances in Neural Information Processing Systems*, 37:141490–141527, 2024.

Olivier Bousquet, Daniel Kane, and Shay Moran. The optimal approximation factor in density estimation. In *Conference on Learning Theory*, pages 318–341. PMLR, 2019.

Olivier Bousquet, Mark Braverman, Gillat Kol, Klim Efremenko, and Shay Moran. Statistically near-optimal hypothesis selection. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 909–919. IEEE, 2021.

Satyaki Mahalanabis and Daniel Stefankovic. Density estimation in linear time. In Rocco A. Servedio and Tong Zhang, editors, *21st Annual Conference on Learning Theory - COLT 2008, Helsinki, Finland, July 9-12, 2008*, pages 503–512. Omnipress, 2008. URL http://colt2008.cs.helsinki.fi/papers/105-Mahalanabis.pdf.