We thank the reviewers for their valuable and helpful comments. Below we address the comments sequentially.

Reviewer #1: *Correct asymptotic* For the properties considered in the manuscript, even the naive empirical-frequency estimator is sample-optimal in the *large-sample* regime (termed "simple regime" in [25]) where the number of samples n far exceeds the alphabet size k. The interesting regime, addressed in numerous recent publications [12, 14, 22, 24, 26], is where n and k differ by at most a logarithmic factor. In this range, n is sufficiently small that sophisticated techniques can help, yet not too small that nothing can be estimated. Since n and k are given, one can decide whether the naive estimator suffices, or sophisticated estimators are needed. Thank you for suggesting and we will make this clear.

Absence of concrete experiments Thank you for asking about the practicality of logarithmic improvements. While some complexity domains look for exponential improvements, for data collection even constant reduction in the number of samples is significant to practitioners. This is one of several recent papers that show a logarithmic reduction over standard estimators. The manuscript does not show new experiments as its main contribution is to theoretically solidify this improvement by showing that it can be achieved by a unified estimator. Some of the results we get are significant. For example, Theorem 1 on Lipschitz property estimation is the most general result we know on this topic.

Specific comments: *Explicit Lipschitz constant in Theorem 1* Sure, the constant is clear from the proof, we'll add it.

Comparisons in Table 2 All the results are ready and we will add comparisons. Line 101 and 234: typos corrected.

Line 277: *Number of symbols is at most k, correct?* Correct, it is also at most n here since we consider symbols in X^n. We will rewrite this paragraph to improve its clarity. Thank you for suggesting.

Reviewer #2: *Results are nice but somewhat incremental* While polynomial approximation has been applied to property estimation, the use of piece-wise polynomials is novel and the analysis of the resulting algorithm is nontrivial (e.g., supplement's Section 2.3). Some of the results are also novel and significant. For example, line 120 shows that all Lipschitz properties can be estimated up to an error of ε using $k/(\varepsilon^2 \log k)$ samples, regardless of symmetry.

The privacy part doesn't fit and *compare with the upper bound of [2] in line 154-155* Please view Theorem 5 as the main result and all the other results as its corollaries. For privacy, our contribution is a unified approach, not the known sample complexity bounds. We mentioned, in line 65-66, that [2] derived tight lower and upper bounds. We apologize for making a mistake and not mentioning the upper bounds in [2] again in line 154-155. We will mention them again in line 154 and clarify that these bounds are known. In addition, we have removed constants 2 in 2ε and 2α.

Comments: Line 73: *Distance estimation using the Valiant-Valiant techniques* Sure, we will introduce this result.

Line 93-96: *The approach is not conceptually different* We provide a different view of property estimation that allows us to simplify the proofs and broaden the range of the results. We will extend the paragraph and make our point clearer.

Line 127-133: *Previous works may also lead to high-probability statements* The major contribution here is a unified approach to deriving high-probability estimators. We are not claiming other methods can not achieve these guarantees, instead, we want to demonstrate that our method has many desired attributes. We compared our result with the median trick approach since it is a natural baseline. We will check related works and clarify this point.

Line 140-141: *Subconstant ε for support size* We need this condition only to derive the simple upper bound presented in Table 2. This is not required by our algorithm, which achieves the minimax MSE for support size estimation.

Line 150: *Conditions for KL divergence estimation* These conditions appear in "Minimax rate-optimal estimation of KL divergence between discrete distributions" and some subsequent papers. We will include these references.

Reviewer #3: *Present the estimator in Section 2* Sure, we will include a high-level description of the final algorithm.

The abbreviation PML is not defined We use "PML" to refer to a different (also well-known) method. The definition of this "PML estimator" is actually quite simple and intuitive, and we will include it. Line 81: belongs –> belongs to.

The bounds for power sum not matching The lower and upper bounds in Table 2 match with each other whenever the first term dominates. We believe that the upper bound is not tight and a finer analysis of our estimator may tighten it.

Line 154-155: *Missing a term* We have added the missing term to the expression.

Line 214: *Statement may be incorrect* This statement should be correct. Basically, we view probability as "degrees of belief", and if we always claim $p \in I_j^*$ whenever $\hat{p}_1 \in I_j$, then we will be correct with high probability. Note that this explanation only provides intuition and is not used in the proof. We will think about how to further clarify.

Line 237: *Should be 'bias' instead of 'error'* Yes, we have corrected this. Thank you for suggesting.

The proof for Theorem 3 is missing in the appendix The proof of Theorem 3 is a direct application of Theorem 5. For each property, we need only a few lines to establish the result. We will provide this short proof in the supplement.