We thank the reviewers for the helpful comments, valuable suggestions, and positive feedback.

Reviewer 1: Thank you for the feedback. We are glad you appreciate our contributions towards obtaining a more complete picture of symmetric property estimation. First, we want to reiterate why we think the regime $k \ll \sqrt{n}$ is interesting in theory. We think this is an important regime in large part due to recent advances in PseudoPML and profile entropy. These results provide general approaches for symmetric property estimation with provably better statistical guarantees under various settings. Previously no provably efficient algorithms were known to implement them; our algorithms enable efficient implementation of these results.

As for the practical considerations of the $k \ll \sqrt{n}$ regime and the experiments, you raise a good point that there are other practical algorithms and heuristic approximate PML algorithms that may achieve comparable estimation error (which our experiments in some cases corroborate). Nevertheless, we believe our results are novel and of interest to the NeurIPS community due to their provable efficiency and worst case guarantees. Our experiments were mainly focused on demonstrating that our provable method can work as well as all previous known estimators including the entropy specific estimator [JVHW15]. Now we do not know how well some of these other heuristic and universal methods perform in the worst case; finding worst case distribution for these other approaches which disprove their efficiency or proving that the other heuristic methods do well, is interesting follow up work, but outside the scope of this result.

As for experiments regarding run time, we want to point out that the algorithms we compare to fall in different categories (heuristic [PJW17], property specific [JVHW15], universal approach for constant error regime [VV11b] and provable PML based approaches [ACSS20, current work]). It is not surprising that the algorithms in the first three categories seem to have better run times than provable PML based approaches as they are either fine tuned for the specific property, lack statistical guarantees, or provably work in a limited error regime. To further improve the running time of the provable PML based approaches involves building an efficient practical solver for a specific convex optimization problem, which we believe is an important problem in itself and outside the scope of our work. Within the category of provable PML based approaches (that includes [ACSS20] and our current work), we believe our algorithm is faster in practice for problems like entropy estimation. Since in our algorithm, we implement the PseudoPML approach that requires the computation of an approximate PML distribution (the major computational bottleneck) on a smaller input size, we have a run time advantage relative to [ACSS20], a vanilla PML based approach (a similar phenomenon was empirically observed earlier in [CSS19b]). We will add a more detailed discussion on the run time of different approaches in the final version of the paper.

Reviewer 2: Thank you for your comments. We agree that we could do better in clarifying our technical contribution. Regarding your concern of the relation of this paper to prior work: yes, we heavily use the previous results from [CSS19a] and [ACSS20] but there is a clear delineation between these prior results and our own. Previous results [CSS19a] and [ACSS20] provided a convex relaxation to the PML objective but rounding the fractional solution with the desired guarantee ($\exp(-k \log n)$) was open – this is our contribution. The rounding algorithms provided in the prior work [CSS19a] and [ACSS20] had worst case approximation ratios lower bounded by $\exp(-n^{2/3} \log n)$ and $\exp(-\sqrt{n} \log n)$ respectively. There were several challenges in obtaining an improved rounding algorithm to obtain an $\exp(-k \log n)$ approximation and thereby obtaining such an improved instance based approximation guarantee. We addressed these challenges by understanding the sparsity structure of the convex relaxation (Lemma 4.3) and further providing a novel matrix rounding algorithm (Theorem 4.4) that draws interesting connections to graph theory. Indeed these are the main steps in obtaining factor k instead of $\sqrt{n}$ in the approximation ratio. Further, we provide a more practical and theoretically provable rounding algorithm for the purposes of PseudoPML and show that it performs well in experiments. We will incorporate a more detailed version of this discussion in the paper.

Reviewer 3: Thank you for the feedback. We are glad you appreciate our contributions and find them novel and powerful. To smooth out the presentation, we will add an overview of techniques section in the final version and explain the proof strategy before getting into technical details. Regarding the practical applications of PML and universal estimators in general, we are unaware of direct immediate practical applications. However, the line of work on universal estimators is quite recent; we hope that in the longer term, these results could possibly yield an efficient toolkit for effectively answering a variety of statistical questions in practice, especially for new symmetric properties lacking custom estimators. We believe the work on PML is broadly related to questions on the phenomenon of universality and might have interesting connections with universal sketches that further have interesting practical applications. Exploring these connections and finding practical applications of PML are very interesting questions and directions for future work.