We thank all the reviewers for liking our paper and providing positive and insightful feedback. Below we address the comments and questions regarding our results and writing style.

Reviewer 1: Thank you again for your encouraging comments on the theoretical and experimental results. We too are happy that a single method could perform really well on a variety of learning tasks. We completely agree that approximating PML is an interesting and important problem for further research, and will look into it in the near future.

Reviewer 2: We really appreciate your thorough and insightful comments. We have incorporated all of them in our draft and will submit the new version if the paper gets accepted. Below we present our detailed responses in order.

– Thank you for pointing out that the introduction section put much of its emphasis on the property estimation problem. We are modifying and reorganizing the introduction to improve its presentation. We will also motivate the other two learning tasks: sorted distribution estimation and property testing.

– L.50 (and abstract): We have changed "distribution estimation" to "sorted distribution estimation";

– Abstract and L.49 and throughout; L.70: We have removed the hyphens in "statistical-learning" and "median-trick";

– L.104, L.285: We have modified the reference list and cited the Charikar et al. paper as a single paper [23]. In the submitted version, we had three different citations because the STOC camera-ready version was not available at that time. Since the result was relatively new, we also included the talk to help potential reviewers better understand it.

– L.115 and L.123: We have removed citation [21] and modified $\log|\epsilon|$ to $\log(1/\epsilon)$.

*** L.130: The constant c used for APML is actually (slightly) worse than that for PML. On the other hand, this makes it possible to strengthen the error probability bound. We have updated our draft to clarify this.

*** Theorem 6: The current proof does not yield the $\sqrt{k}/\varepsilon^2$ complexity of uniformity testing in the constant confidence regime. We do have an alternative argument that utilizes the problem structure to achieve the $\sqrt{k}/\varepsilon^2$ sample complexity. We will provide a sketch of the alternative argument in the updated version.

– L.165: The emphasis here is that our tester is "the first PML-based uniformity tester" providing both the ℓ_1 and ℓ_2 testing guarantees. Incorporating your comments, we have modified the statement and pointed out that "nearly all uniformity testers in the literature [...] provide the same ℓ_2 testing guarantee".

– L.251: We have added a short motivation for the sorted ℓ_1 distance estimation. There are several motivations, including but not limited to (unsorted) distribution estimation [70] and symmetric property estimation.

– L.271: We have removed [3] and [17]. We appreciate the detailed comments regarding the references.

"Question": A simple definition of a *universal plug-in property estimator* could be a sentence similar to the one in the abstract. For example, "there exist absolute positive constants c_1, c_2 and c_3, such that for any 1-Lipschitz property on $(\Delta_{\mathcal{X}}, R)$, with probability $\geq 9/10$, the plug-in estimator uses just c_1 times the sample size n required by the minimax estimator to achieve c_2 times its error, whenever this error is at least n^{-c_3}". We are still thinking about better definitions.

Reviewer 3: Thank you for the encouraging comments. The concise and detailed summary of the paper's contributions you provided is valuable for us to improve its presentation and organization. We are also excited about the broad optimality of PML and APML as well as your strong recommendation for the paper's acceptance. Below we mainly address the "Cons" mentioned in your comments.

1. We agree that the property testing result is not as impressive as the paper's other results. Yet we still find it interesting because combined with other results, it shows that PML is a generic tool for a variety of inference tasks. In addition, the $\exp(-3\sqrt{n})$ error probability bound does not hold for property testing [31], hence the "competitiveness arguments" in [5] do not directly apply here. Instead, we utilize a new concept of "typical profiles" to reduce the number of objects considered, which in turn requires only weaker concentration. This technique may be of independent interest.

2. Thank you for the nice suggestion of performing experiments using APML, the near-linear computable variant of PML. The code is not publicly available yet, and we will ask the authors of [23] for the code so we can compare APML and the MCMC-EM algorithm. We are looking forward to seeing the experimental results.

3. The experiments for Shannon entropy estimation basically showed that the MCMC-EM PML computation algorithm is as good as state-of-the-art algorithms specifically designed for entropy estimation. One way to improve the estimation accuracy is to increase the number of EM iterations, e.g., from 30 to 50. On the other hand, this will make the algorithm around two times slower, since the computation time of each MCMC-EM iteration is roughly the same.

4. Detailed comments: 1) We have modified the notation used in line 78-79 according to your suggestion; 2) We have simplified the statements of conditions and made Theorem 1 self-contained; 3) We have added a new paragraph to give a high-level description of the APML algorithm. Thank you for helping us enhance the writing of the paper.