We thank all reviewers for their valuable comments and feedback. Please see replies to comments below.

Reviewer 1

High probability bounds: We agree with the reviewer that obtaining a bound with dependence on the confidence parameter is a very interesting problem. Our algorithm guarantees (upper bounds) can be extended to high probability bounds with a $\log(1/\beta)$ dependence, where β is the error probability. We provided guarantees in expectation as it makes expressing min-max rates and obtaining lower bounds easier. If accepted, we will add a discussion regarding the high probability bounds in the final version.

Reviewer 2

We thank the reviewer for the comments. We will highlight the specific novelties such as restricted estimators and bounds on total variation distance between binomial distributions in the final version.

Reviewer 3

We thank the reviewer for suggestions on the writing including adding pseudo-code for the algorithms. If accepted, we will incorporate them in the final version.

I.i.d. data: Distributed learning of discrete distributions when samples are generated from a single same distribution has been studied extensively with traditional item-level differential privacy including Duchi et al. (2013), Kairouz et al. (2016) (local differential privacy), and Diakonikolas et al. (2015), Acharya et al. (2020) (global differential privacy). It is also common in communication constrained settings such as Barnes et al. (2019).

Our work extends these results into user-level privacy and is the first step towards understanding utility-privacy trade-offs in user-level privacy. We believe such an analysis would give insights to design algorithms that might perform well in the non i.i.d. setting. We agree that extending the results to distinct distributions is an interesting future direction.

Different number of samples per user: We have proposed a modified algorithm for the case when users have different number of examples (see lines 141-143). Due to space constraints, the details are described in Appendix E. The user complexity is similar to the case when all users have m samples, with m replaced by the median of number of user samples. We will highlight this in the main paper.

Finer instance-specific bounds and other metrics: We agree that these problems are interesting and would explore them in future works.

Nodal differential privacy: We thank the reviewer for the reference. We will add a discussion on it and explore future work in this direction.

Lemma 3: We are inverting the function $y = (1-p)^m$ to compute the p. Such an estimator is only statistically efficient when p is small. This is quantified by an upper bound on p, given by c/m. Hence, we apply this subroutine for $p < c/m$, where c can be as small as 2 or 3.

Reviewer 4

We thank the reviewer for the positive comments.

I.i.d. data: Distributed learning of discrete distributions when samples are generated from a single distribution has been studied extensively with traditional item-level differential privacy extensively including Duchi et al. (2013), Diakonikolas et al. (2015), Kairouz et al. (2016), and Acharya et al. (2020). Our work extends these results into user-level privacy and is the first step towards understanding utility-privacy tradeoffs in user-level privacy. We agree that extending the results to distinct distributions is an interesting future research direction. We finally note that learning discrete distributions with user-level DP is an important practical problem for applications such as word prediction in virtual mobile keyboards (with/without federated learning).

- Duchi et al. (2013): Local privacy and statistical minimax rates.
- Diakonikolas et al. (2015): Differentially private learning of structured discrete distributions
- Kairouz et al. (2016): Discrete distribution estimation under local privacy
- Acharya et al. (2020): Differentially private assouad, fano, and le cam.
- Barnes et al. (2019): Lower bounds for learning distributions under communication constraints via fisher information.