We want to thank all the reviewers for their insightful and valuable comments. All the typos have been corrected. The rest of the rebuttal addresses specific comments and questions about our results and writing style.

Reviewer 1: [*Storyline*] Thanks for commenting on the paper's writing style. Our paper proposes a very novel quantity and aims to derive a thorough theory for it. Hence, the writing style differs from those that solve a particular problem.

[*Compression and Adaptive Algorithms*] We stated the block compression algorithm just one line below the content you referred to – in line 227, "we can compress $\varphi(x^n)$ into the set $\mathcal{C}(\varphi(x^n))$ of corresponding multiplicity-prevalence pairs." The sentence in line 22, in its context, refers to the generic action of designing an adaptive algorithm for the problem at hand. We have replaced "we design" by "one can design" to eliminate ambiguity.

[*Method of Types*] The word "connects" means "establishing a link" instead of "precisely characterizing the relation." We agree that the relation between the profiles and the method of types is worth studying. On the other hand, the current paper already contains 8+ theorems (a few more in the supplementary), 40+ pages proof, and multiple program modules. Hence, we presented this connection as a future research direction (extension) instead of a major result in line 333-338.

[*Old and New Results*] Since the paper addresses three problems and presents eight theorems, we have, as stated in line 60, "relegate(d) detailed reviews on related work . . . to the supplementary material." As a result, we have assumed that the readers are somewhat familiar with the related work. If not, we expect the readers to look at the supplementary material, which is a good source for background knowledge. An example is property estimation, for which Appendix B devotes 1.5 pages in B.3 to discuss the prior results and methods. Thanks for your suggestion on presenting such background materials in the main paper. We plan to add shorter versions of the prior-work reviews in the supplementary to the beginning of each of the five subsections of Section 2. This should be relatively easy because most theorems are either entirely new (e.g., thms. 1, 3, 5, and cors. 1, 2) or already equipped with the essential references (e.g., thms. 2, 4).

[*Clarity*] Thanks for suggesting compressing the first part of Section 2.4 and including it in the introduction. We will modify the paper accordingly, and also reduce the usage of the \paragraph{ } command.

[*Example and Notation*] We will add an example for the sample profile and its entropy in Section 1. For the notation, φ_μ is a function, while Φ^n is a random multiset, and we followed the convention of capitalizing the random quantities; $\mathcal{D}_n \simeq \mathcal{H}_n$ means that for any n, the two quantities are of the same magnitude, up to a logarithmic factor of n (line 44).

Reviewer 2: [*Implications for Practical Problems*] Thanks for the encouraging comments and for suggesting providing insights on practical applications. The problems we studied in this paper – distribution estimation, property inference, and profile compression, are quite fundamental. They have wide applications to numerous disciplines, as outlined in line 79-81, 140-143, and 192-210. Hence, we did not further emphasize the practical implications.

[*Profile Entropy Computation*] For high-dimensional alphabets, even computing a single profile probability often takes exponential time, and there can be $\exp(\Theta(\sqrt{n}))$ such probabilities. Hence, computing the profile entropy $\mathcal{H}_n$ can be computationally expensive. On the other hand, Theorem 1 shows that $\mathcal{D}_n$, a quantity computable in near-linear time, nicely approximates $\mathcal{H}_n$. Appendix A.4 further illustrates how to estimate $\mathcal{H}_n$ with $m \ll n$ observations.

[*Seeing the PML Estimator*] Thanks for your interest in PML. We do not quite understand the meaning of "see the PML" here. In case you refer to computing the PML precisely, the paper "algebraic computation of pattern maximum likelihood (ISIT 2011)" provides an exponential-time algorithm with concrete examples. On the other hand, researchers often employ algorithms that approximate PML, such as the polynomial-time algorithm mentioned in line 183.

[*Asymptotic Closeness between $\mathcal{D}_n$ and $\mathcal{H}_n$*] As both quantities highly depend on the underlying distribution and can be as large as $\Theta(\sqrt{n})$, currently we do not see the possibility of characterizing the general asymptotic relation.

[*Connection to the Method of Types*] Please see the third point in our response to Reviewer 1 – [*Method of Types*].

[*Empirical Analysis*] Thanks for this point. Please note that we do have 2 sets of experiments in Appendices B.5 and C.4 with complete code accompanied. We did not test the compression algorithm as it was proven to be instance-optimal.

[*Connection to Universal Source Coding*] Thanks for suggesting. Sure, we'll add a section to illustrate this connection.

Reviewer 3: Thanks for the insightful suggestions and valuable comments.

[Optimality of PML] We agree that determining the scope of PML's optimality is an important problem. Though Theorem 1 does not recover some of the prior results on the PML, it applies to all symmetric properties and distribution collections, hence covering a much broader class of property inference problems.

[Upper Bound in Theorem 2] To be absolutely clear, we have included the statement in line 113 as part of the theorem.

[Lower Bound Statement of Theorem 2] Excellent suggestion on how to present the lower bound. The suggested form is mathematically equivalent to the original but is more compact. We have modified the theorem's statement accordingly.