Estimating the True Distribution of Data Collected with Randomized Response

Carlos Antonio Pinzón[1], Ehab ElSalamouny[1,5], Lucas Massot[3], Alexis Miller[4], Héber Hwang Arcolezi[2], Catuscia Palamidessi[1]

[1]INRIA Saclay, France
[2]INRIA Grenoble, France
[3]École Polytechnique, France
[4]Ecole Normale Supérieure de Lyon, France
[5]Suez Canal University, Egypt

Abstract

Randomized Response (RR) is a protocol designed to collect and analyze categorical data with local differential privacy guarantees. It has been used as a building block of mechanisms deployed by Big tech companies to collect app or web users' data. Each user reports an automatic random alteration of their true value to the analytics server, which then estimates the histogram of the true unseen values of all users using a debiasing rule to compensate for the added randomness. A known issue is that the standard debiasing rule can yield a vector with negative values (which can not be interpreted as a histogram), and there is no consensus on the best fix. An elegant but slow solution is the Iterative Bayesian Update algorithm (IBU), which converges to the Maximum Likelihood Estimate (MLE) as the number of iterations goes to infinity. This paper bypasses IBU by providing a simple formula for the exact MLE of RR and compares it with other estimation methods experimentally to help practitioners decide which one to use.

1 Introduction

Local differential privacy (LDP) (Duchi, Jordan, and Wainwright 2013) is a framework for providing privacy guarantees when collecting data from a set of users. It removes the need for trust in the correct management of whoever collects the data and is typically used for automated telemetry. The essence of LDP is to introduce controlled uncertainty right before the users transmit their data to avoid the data collector from receiving the exact specific information of each user.

The level of uncertainty or noise introduced by LDP is controlled by a parameter $\epsilon > 0$, varying from a high privacy regime with $\epsilon \approx 0$ (large uncertainty) to a low privacy regime with $\epsilon \gg 1$ (low uncertainty). The organization that collects and processes the data can not be certain about the specifics of any user in particular, but it can still derive statistical estimations. These estimations become more precise as ϵ increases and as the size of the user group grows.

One of the simplest and most fundamental LDP mechanisms is the Randomized Response (Kairouz, Bonawitz, and Ramage 2016) (also called K-ary RR, kRR, or simply RR), introduced firstly without any connection to LDP in (Warner 1965). In a nutshell, RR reports either the truth with some fixed probability or a random value, picked uniformly.

Method	Complexity	Justification
Debiasing `Inv`	$O(K)$	✓ Unbiased, but invalid
Neg. fix `InvN`	$O(K)$	Workaround: simple
Proj. fix `InvP`	$O(K \log K)$	Workaround: close to `Inv`
Bayesian `IBU`	$O(K\, N_{\text{iters}})$	✓ MLE when $N_{\text{iters}} \to \infty$
Proposed `MLE*`	$O(K \log K)$	✓ MLE in one step

Table 1: Methods for estimating the original distribution from RR observations. N_{iters} represents the number of iterations, which depends on the desired precision.

Despite its simplicity, RR has been the subject of extensive research as it is the building block of more advanced mechanisms like Symmetric Unary Encoding and Local Hashing (Wang et al. 2017) as well as longitudinal protocols, e.g. RAPPOR (Erlingsson, Pihur, and Korolova 2014), d-bitFlipPM (Ding, Kulkarni, and Yekhanin 2017), Longitudinal Local Hashing (Arcolezi et al. 2022).

Moreover, RR is known to be optimal among all LDP mechanisms for small domains or large values of ϵ. Namely, whenever the domain size K satisfies $K < 3e^{\epsilon} + 2$ (Wang et al. 2017). Large values of ϵ are particularly relevant in light of recent advances in Shuffle differential privacy (Erlingsson et al. 2019; Cheu et al. 2019), where a shuffler is applied after local randomization to strip metadata and randomly permute the reports. This added layer of anonymity enables privacy amplification, meaning that stronger central-DP guarantees can be achieved from weaker local ones. The growing use of Shuffle DP in real-world scenarios heightens the importance of high ϵ regimes, and therefore, of RR.

Concerning statistical estimations, we focus on estimating the original data distribution, which is arguably the most fundamental statistic. To this purpose, typically the analytics server constructs a histogram of reported data, normalizes it to obtain a distribution, and uses a debiasing linear correction rule (`Inv`) to compensate for the added noise. The resulting vector is proven to be unbiased, but it may not be a valid distribution, as it may have some negative entries. In order to obtain a distribution, two simple workarounds have

been proposed. The first, InvN, sets to 0 all the negative entries of the vectors, and renormalizes it. The second, InvP, projects the vector on the closest point on the simplex in terms of Euclidean distance.

A different approach, called iterative Bayesian update (IBU), was proposed by (Agrawal and Aggarwal 2001). IBU produces an estimate of the original distribution using an iterative algorithm that converges to the maximum likelihood estimate (MLE) as the number of iterations grows (El-Salamouny and Palamidessi 2020). In this sense, IBU is well justified mathematically. However, the asymptotic nature of the algorithm makes it less attractive. Another drawback is the missing theory about the number of iterations or the stop condition that guarantees a certain proximity to the MLE.

Our first objective is to enhance the efficiency of MLE computation. To this aim, we propose MLE$\star$, based on a mathematical formula that can be computed quickly, and which provides the exact MLE, thus bypassing the issue about the stop condition.

The second objective is to compare these estimators. One may think that, since the MLE is the "most likely" estimate, it would also be the most precise. Furthermore, the MLE is known to minimize the Kullback-Leibler divergence between the empirical distribution (i.e., the normalized histogram output by RR) and the ideal distribution, obtained by multiplying the input distribution by the channel matrix that represents the RR noise. However, for practical purposes, the Kullback-Leibler divergence may not be the most important utility metric. Indeed, the precision of an estimation is typically measured in terms of mean square error (MSE). In this paper, we compare the MSE precision of the MLE, InvN and InvP. It turns out that, surprisingly, the precision of InvN and InvP "flip" depending on the distribution being more or less concentrated, while that of the MLE is always in between.

Other important terms of comparisons we investigate in this paper are efficiency, consistency, and unbiasedness. The methods that we consider are summarized in Table 1.

In summary, the contributions of this paper are:

1. We provide a simple formulation for the MLE, prove its correctness formally, and validate it empirically. These formulas coincide with (Kairouz, Bonawitz, and Ramage 2016), Supplementary Material, Section F.
2. We propose an algorithm that computes the MLE and is significantly more efficient than the state-of-the-art solution (IBU). In addition, our algorithm gives an exact solution, whereas IBU provides only an approximation because the number of iterations is necessarily finite.
3. We theoretically and experimentally compare MLE with other estimation methods to clarify their trade-offs, with the objective to help developers choose which one to use.

2 Preliminaries

This section formulates the estimation problem, presents the state-of-the-art estimators from related work.

2.1 Problem formulation

There are N users $u = 1..N$, each of which has a secret value x_u from a set of K categories labeled as $\{1, 2, \ldots, K\}$. The secret value can represent, for example, the user's browser homepage, the number of times they click on a specific button, or the emoji they use most frequently.

An organization, referred to as the data collector or the analyst, is interested in finding out the most frequent categories of the population as a whole (not user by user), and more generally, they are interested in estimating the proportion θ_i of users u taking value $x_u = i$ for every $i \in \{1..K\}$. This corresponds to finding a vector $\theta \in \Delta$ where the simplex Δ is the space of distributions over $\{1..K\}$, i.e., $\Delta := \{\theta \in \mathbb{R}^K : \sum_i \theta_i = 1, \theta_i \geq 0\}$.

To avoid violating the privacy of the users, the data collector will not collect x_u directly. Instead, they will use a *mechanism* $\mathcal{M}$ (a function whose output is influenced by randomness, also known as a channel in information theory), whose purpose is to deliberately transform the input data x_u into some $y_u := \mathcal{M}(x_u)$ with controlled randomness to hide its true value while still providing some information about it. One sample per user is measured.

We suppose that $\mathcal{M}$ is an instance of RR, and address the problem of how the analyst efficiently estimates the unknown distribution $\theta \in \Delta$ using the outputs $(y_u)_{u=1}^N$ and the structure of the mechanism $\mathcal{M}$.

2.2 LDP and the RR mechanism

Local Differential Privacy (LDP) (Duchi, Jordan, and Wainwright 2013) is a strong privacy protection ensured by the user-side mechanism by setting a formal bound on how much information is revealed about the true input. Formally, a mechanism $\mathcal{M}$ with discrete output space satisfies ϵ-LDP for some $\epsilon > 0$ (the smaller the more private) if for all inputs x, x' and outputs y, it holds that

$$\Pr(\mathcal{M}(x) = y) \leq e^{\epsilon} \Pr(\mathcal{M}(x') = y).$$

This protects the unknown input because upon observing any output y, the degree to which any candidate input x is more likely than any other x' is limited by ϵ.

The randomized response (RR) mechanism takes an input $x \in \{1..K\}$ and reports $\mathcal{M}(x)$ from the same domain such that

$$\Pr(\mathcal{M}(x) = y) = \begin{cases} p & \text{if } y = x \\ q = \frac{1-p}{K-1} & \text{otherwise} \end{cases} \quad (1)$$

That is, $\mathcal{M}(x)$ returns x with probablility p, and otherwise outputs some $x' \sim \text{uniform}(\{1, ..., K\} \setminus \{x\})$. Equivalently, it returns x with probability $p - q$, and otherwise returns some $x' \sim \text{uniform}(\{1, ..., K\})$. The RR was first introduced by Warner (1965) for the binary case ($K = 2$), as a survey technique for eliminating evasive answer bias. More recently, RR has gained increased attention in the context of LDP, since it satisfies ϵ-LDP for $\epsilon = \log(p/q)$.

2.3 Estimators based on expectation

Let $\phi \in \Delta$ denote the normalized histogram of the observed y_u for $u = 1..K$, after applying RR parametrized by some

known value p. Note that the expected value of ϕ is $\phi_i = q+(p-q)\theta_i$. Based on this observation, Kairouz, Bonawitz, and Ramage (2016) derived a simple and unbiased estimator, which we call the *linear inversion* estimator (`Inv`)

$$\hat{\theta}_i^{\texttt{Inv}} := \texttt{Inv}(\phi)_i := \frac{\phi_i - q}{p - q}. \tag{2}$$

Nevertheless, the constraint $\hat{\theta}^{\texttt{Inv}} \in \Delta$ may fail to hold, since (2) may produce negative values. This occurs very often, as detailed later in Section 5.1. To overcome this issue, the following two post-processing solutions have been proposed in the literature.

1. Normalization (`InvN`): $\hat{\theta}_i^{\texttt{InvN}} \propto \max(0, \hat{\theta}_i^{\texttt{Inv}})$, i.e., set to zero the negative components and rescale.
2. Projection (`InvP`): $\hat{\theta}^{\texttt{InvP}} := \arg\min_\theta \|\theta - \hat{\theta}^{\texttt{Inv}}\|$, where θ varies in the set of valid distributions.

These estimators are valid (they always produce distributions), and they are fast. `InvN` is (clearly) $O(K)$ and `InvP` can be implemented in $O(K \log K)$. However, their design is a post-processing workaround without formal guarantees.

3 Estimators Based on Likelihood

Let $\mathbf{MLE}(\phi)$ be the set of all MLEs. That is

$$\mathbf{MLE}(\phi) := \overset{\text{(set)}}{\arg\max_{\theta \in \Delta}} \Pr(\phi|\theta), \tag{3}$$

where $\Pr(\phi|\theta)$ denotes the probability that the normalized histogram of the random variables Y_u for $u = 1..N$, is ϕ provided that the distribution for x_u is θ and $Y_u \sim \mathcal{M}(x_u)$. Based on the theory of the Expectation-Maximization algorithm (Dempster, Laird, and Rubin 1977), it has been shown (Agrawal and Aggarwal 2001; Agrawal, Srikant, and Thomas 2005; ElSalamouny and Palamidessi 2020) that one can asymptotically approach an MLE using an algorithm known as Iterative Bayesian Update (IBU). This algorithm is applicable to any discrete mechanism with a finite channel matrix $C_{i\,j} := \Pr(\mathcal{M}(i) = j)$ of size $K_{\text{in}} \times K_{\text{out}}$.

IBU starts with a fully supported prior distribution $\hat{\theta}^{(0)}$, by default $\hat{\theta}^{(0)} := (1/K_{\text{in}}, 1/K_{\text{in}}, ..., 1/K_{\text{in}})$, and repeatedly updates $\hat{\theta}^{(t)}$ into $\hat{\theta}^{(t+1)}$ using what is known as Jeffrey's update rule (Jacobs 2021; Pinzón and Palamidessi 2025):

$$\hat{\theta}_i^{(t+1)} := \sum_{j=1}^{K_{\text{out}}} \frac{\phi_j\, \hat{\theta}_i^{(t)} C_{i\,j}}{\sum_{k=1}^{K_{\text{in}}} \hat{\theta}_k^{(t)} C_{k\,j}}. \tag{4}$$

The most important property of IBU is that $\hat{\theta}^{(t)}$ converges to $\hat{\theta}^*$ for some $\hat{\theta}^* \in \mathbf{MLE}(\phi)$ as $t \to \infty$ (ElSalamouny and Palamidessi 2020). The complexity of this procedure for a fixed large number of iterations $t = N_{\text{iters}}$ is $O(K_{\text{in}}\, K_{\text{out}}\, N_{\text{iters}})$, because at each time step t, all denominators (for every j) can be cached in $O(K_{\text{in}}\, K_{\text{out}})$ and then $\hat{\theta}^{(t+1)}$ can be computed in $O(K_{\text{in}}\, K_{\text{out}})$.

In the particular case of RR, where $K_{\text{in}} = K_{\text{out}} = K$, the algorithm can be accelerated from $O(K^2\, N_{\text{iters}})$ to $O(K\, N_{\text{iters}})$ using symmetries. Precisely, the update step (4) can be simplified to the following, which takes only $O(K)$.

$$\begin{aligned} s^{(t+1)} &:= \sum_{i=1}^{K} \frac{\phi_i}{q + (p-q)\, \hat{\theta}_i^{(t)}} \\ \hat{\theta}_i^{(t+1)} &:= \hat{\theta}_i^{(t)} \left(q\, s^{(t+1)} + \frac{(p-q)\, \phi_i}{q + (p-q)\, \hat{\theta}_i^{(t)}} \right). \end{aligned} \tag{5}$$

We will denote this procedure by $\texttt{IBU}(\phi) := \hat{\theta}^{(N_{\text{iters}})}$ for some fixed N_{iters}.

Lastly, Ye et al. (2025) propose an algorithm that optimizes a regularized version of the likelihood by merging small values together. Their algorithm runs in $O(K^2 \log(K)\, N_{\text{iters}})$ and uses a smoothing factor, which, for some practical experiments, gives better results than a pure MLE. Hay et al. (2009) address the estimation under central (not local) differential privacy, and Lee, Wang, and Kifer (2015) propose an approximate iterative method (like IBU).

4 Formula for the MLE

This section starts with a derivation of the formula for the MLE, which coincides in different notation with that of (Kairouz, Bonawitz, and Ramage 2016), Supplementary Material, Section F, and then provides the pseudocode to compute it and a summarizing sketch of the detailed self-contained proof in the Supplementary Material.

Consider a RR mechanism, defined with $p > 0$ and $q = \frac{1-p}{K-1}$ as in (1). For any observed distribution ϕ and some threshold $\tau \in [\min_i \phi_i, \max_i \phi_i]$, define ϕ^τ as

$$\phi_i^\tau = \begin{cases} q & \text{if } \phi_i < \tau \\ c_\tau\, \phi_i & \text{otherwise,} \end{cases} \qquad c_\tau = \frac{1 - \sum_{\phi_i < \tau} q}{\sum_{\phi_i \geq \tau} \phi_i}.$$

Notice that $\sum_i \phi_i^\tau = 1$ by the definition of the constant c_τ. Observe also that the denominator in c_τ is non-zero because $\tau \leq \max_i \phi_i$. For this threshold transformation, notice that

$$\texttt{Inv}(\phi^\tau)_i = \begin{cases} 0 & \text{if } \phi_i < \tau \\ \frac{c_\tau \phi_i - q}{p-q} & \text{otherwise.} \end{cases} \tag{6}$$

The vector $\texttt{Inv}(\phi^\tau)$ sums up to 1, but it may contain negative values for small τ. The proposed estimator, which we call $\texttt{MLE}\star(\phi)$, is precisely $\texttt{Inv}(\phi^{\tau^*})$ where τ^* is the smallest threshold τ for which $\texttt{Inv}(\phi^\tau)$ does not contain negative values:

$$\begin{aligned} &\hat{\theta}^{\texttt{MLE}\star} := \texttt{MLE}\star(\phi) := \texttt{Inv}(\phi^{\tau^*}), \\ &\tau^* := \min\{\tau \mid \forall i,\ \phi_i < \tau \vee\ c_\tau\, \phi_i \geq q\}. \end{aligned} \tag{7}$$

Algorithm 1 computes $\hat{\theta}^{\texttt{MLE}\star}$ for 1-indexed arrays, and we give a zero-indexed Python implementation in the Supplementary Material. The following invariant holds during the loop: s is the suffix sum $s = \sum_{i>k} \phi_{\sigma(i)}$. After the loop, i is the number of zeros in the output $\hat{\theta}^{\texttt{MLE}\star}$ and $\tau^* = \phi_{\sigma(i)}$. The main loop that keeps track of s and finds the threshold τ^* is $O(K)$, therefore, the algorithm as a whole is $O(K \log K)$ because it sorts the indices of ϕ before entering the main loop.

Algorithm 1 Proposed algorithm MLE$\star$. Indexed from 1.

Input: $\phi_{1..K}$, p, q.
$\sigma \leftarrow \text{argsort}(\phi_{1..K})$ ▷ $\phi_{\sigma(1)} \leq \cdots \leq \phi_{\sigma(K)}$
$i \leftarrow 0, \quad s \leftarrow 1$
while $i < K$ and $qs > \phi_{\sigma(1+i)}\,(1 - iq)$ **do**
    $s \leftarrow s - \phi_{\sigma(1+i)}, \quad i \leftarrow i + 1$
end while
$\hat{\theta}^*_{\sigma(j)} \leftarrow 0$ for each $j = 1, ..., i$.
$\hat{\theta}^*_{\sigma(j)} \leftarrow \frac{\phi_{\sigma(j)}\,(1-iq)-sq}{s(p-q)}$ for each $j = i{+}1, ..., K$.
return $\hat{\theta}^*_{1..K}$

We conclude this section by showing that MLE$\star$ is the unique MLE of the RR mechanism, and we show a theoretical application of this result.

Theorem 1. *The MLE for the RR mechanism is unique and given by the proposed formula MLE$\star$.*

Proof sketch. The complete proof is given in Theorem 7 (Supplementary Material). The following is a sketch of it.

Suppose that $\hat{\theta}$ is an MLE. It can be shown using the Lagrange Multipliers method that for every category i, either $\hat{\theta}_i = 0$ or $\hat{\theta}_i = \frac{\phi_i}{\lambda} - \frac{q}{p-q}$, where the value of λ is such that the total sum is 1, i.e. $\lambda := \frac{(p-q)\sum_{\hat{\theta}_i>0}\phi_i}{1-|\{i:\hat{\theta}_i=0\}|q}$. This is proven in Theorem 4 (Supplementary Material). With this result, the problem of computing all the components of $\hat{\theta}$ is reduced into identifying the components i in which $\hat{\theta}_i = 0$ because λ and all the other components can be computed with the given formulas. The rest of the proof is devoted to identifying the zero-valued components.

As shown in Theorem 6 (Supplementary Material), the entries in $\hat{\theta}$ are monotonic with respect to the entries in ϕ in the sense that $\hat{\theta}_i \leq \hat{\theta}_j$ if and only if $\phi_i \leq \phi_j$. This observation follows by contradiction: if $\hat{\theta}$ was not monotonic with respect to ϕ, then the order of two entries can be flipped, resulting in a new $\hat{\theta}'$ that has a higher likelihood.

Therefore, the zeros of $\hat{\theta}$ must occur in the positions of the smallest n components of ϕ. In other words, letting $\theta^{[n]}$ be the result of applying the aforementioned formula obtained via Lagrange Multipliers assuming that the zeros occur in the positions of the smallest n values of ϕ, then the MLE must be $\theta^{[n]}$ for some $n \in \{0, ..., K-1\}$ whose value is unknown so far.

Finally, to find the value of n, let e_i denote the value of the i'th smallest component of ϕ and define $g(n) := (1 - nq)e_{n+1} - q(\sum_{i>n} e_i)$ for $n < K$, then the desired n must satisfy $g(n) \geq 0$ and must be as small as possible. This is shown in Lemmas 3 and 4 (Supplementary material). The threshold τ^* in Equation (7) corresponds to e_{n+1}.

In summary, $\hat{\theta} = \theta^{[n]}$, so the MLE is unique. □

Thanks to Theorem 1, it is possible to study and prove analytical properties of the MLE, like the following, which partly explains some of the results in Section 6.2.

Theorem 2. *If the smallest entry in ϕ, say ϕ_i, takes value $\phi_i < q$, and the second smallest is at least $(Kq - \phi_i)/(K-1)$, then the points in $\mathbb{R}^K$ given by $\hat{\theta}^{InvP}$, $\hat{\theta}^{MLE\star}$ and $\hat{\theta}^{InvN}$ are collinear, with $\hat{\theta}^{MLE\star}$ in the middle.*

Proof. Proven in Theorem 8 (Supplementary Material). □

5 Theoretical Comparison

This section compares validity, unbiasedness, MSE, consistency and complexity for the estimators of interest: MLE$\star$, Inv, InvN and InvP.

5.1 Validity

By validity, we denote the property that the returned estimates are always guaranteed to be valid distributions. Inv guarantees $\sum_i \hat{\theta}_i = 1$ but not $\hat{\theta}_i \geq 0$, therefore, it is invalid. The other three estimators are valid by design.

The problem that Inv produces vectors with negative values occurs very frequently in practice. Indeed, for a population with $\theta_i \approx 0$ for some i, the probability of $\hat{\theta}_i < 0$ equals that of $\phi_i < q$, which is mostly governed by the probability of a Binomial$(N-1, q)$ not exceeding Nq. This value is close to $1/2$ and therefore non-negligible regardless of N.

5.2 Unbiasedness

It was shown in Wang et al. (2017), Theorems 1 and 2, that Inv is unbiased and has element-wise variance given by

$$\text{Var}(\hat{\theta}_i) = \frac{q(1-q)}{N(p-q)} + \theta_i \frac{(p-q)(1-2q-(p-q))}{N(p-q)}, \quad (8)$$

which follows from the observation that ϕ_i follows a Multinomial distribution.

InvN, InvP and MLE$\star$, however, are biased for certain θ, and this is an inevitable consequence of being valid. For instance, fix $\theta = (1, 0, ..., 0)$ and vary ϕ. The realizations of the random vectors ϕ and $\hat{\theta}$ are points scattered around $\phi' = ((p-q), q, ..., q)$ and θ respectively. Some of the points around θ are outside the valid region and some are inside, but they balance and we have $E(\hat{\theta}) = \theta$. If we apply any rule f that moves invalid points to the valid region and keeps valid points as they are, then $E(f(\hat{\theta}))$ will necessarily fall in the (strict) interior of the valid region. Thus, since θ is in the border, not in the interior, we obtain $E(f(\hat{\theta})) \neq \theta$. A generic proof for this fact can be found in Berger (1990).

5.3 Complexity

In terms of computational speed, Inv, InvN and InvP are $O(K)$ while IBU is $O(K\,N_{\text{iters}})$ and MLE$\star$ is $O(K \log K)$ (cost of sorting the indices). The larger cost of MLE$\star$ above $O(K)$ is very low relative to the gained guarantees: the output is valid, it produces the exact MLE, and the complexity difference is sub-polynomial.

5.4 MSE, TV and consistency

The Mean Squared Error is defined as $\text{MSE}_\theta(\hat{\theta}) := E_{\phi|\theta}(\|\hat{\theta} - \theta\|^2)$, where $\hat{\theta}$ is fixed to one of the four estimators. As shown in Theorem 9 (Supplementary Material), the four estimators are consistent because $\text{MSE}_\theta(\hat{\theta}) \in O(K/N)$, which converges to 0 as $N \to \infty$. This implies that other measures, like the Total Variation $\text{TV}_\theta(\hat{\theta}) := \frac{1}{2}E_{\phi|\theta}(\|\hat{\theta} - \theta\|_1)$ also converge to 0 as $N \to \infty$. Alternatively, the consistency of MLE$\star$ can be proven using the sufficient conditions derived by ElSalamouny and Palamidessi (2025) for the consistency of MLE for generic privacy mechanisms, of which RR is a particular case.

6 Experiments

In this section, we describe our experimental setup, present key results, and discuss the implications of our findings. The code is available at (Pinzón et al. 2025). Our experiments pursue two main objectives: **(1)** validating the correctness of our MLE$\star$ estimator by comparing it against the IBU approach under the RR mechanism, and **(2)** evaluating the performance and robustness of MLE$\star$ compared to two standard baselines, InvP and InvN. Specifically, while InvP and InvN perform well in distribution-specific settings, MLE$\star$ offers a consistently strong performance across a wide range of scenarios, making it a robust and reliable choice when the true data distribution is unknown.

6.1 General Setup

We conduct an extensive empirical evaluation to compare the performance of three estimators under the RR mechanism: MLE$\star$, InvP, and InvN.

Experiments are designed to assess robustness across a wide range of configurations. Specifically, we vary:

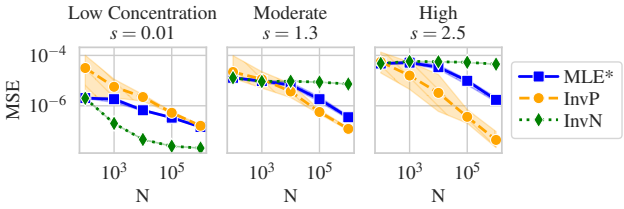
- Privacy budget $\epsilon \in \{1, 2, \ldots, 9, 10\}$;
- Sample size $N \in \{10^2, 10^3, 10^4, 10^5, 10^6\}$;
- Domain size $K \in \{50, 100, 1000, 5000\}$;
- Data skewness, controlled via a Zipf parameter s:
  - Low concentration (near-uniform): $s = 0.01$,
  - Moderate concentration: $s = 1.3$,
  - High concentration: $s = 2.5$.

Each configuration is repeated with 100 different random seeds to account for statistical variability, and the performance is evaluated using: (i) Mean Squared Error (MSE), and (ii) Negative Log-Likelihood.

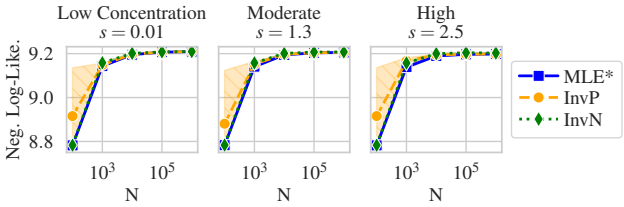
6.2 Overview of Results

To support a comprehensive evaluation, we conduct an extensive set of experiments varying the key parameters of Section 6.1. All experimental results, covering the full grid of configurations, are provided in the supplementary material (Section D). In this section, we highlight representative scenarios that illustrate the main trends and insights, focusing on how each estimator performs under different data regimes and parameter settings.

Evaluation. Figure 1 presents performance results for a practically relevant industrial setting with a large fixed domain size $K = 10{,}000$ and privacy level $\epsilon = 4.0$, while varying the number of users N. This choice of K reflects the high-cardinality domains commonly encountered in real-world applications, such as telemetry data by Google Chrome and Microsoft Windows (Erlingsson, Pihur, and Korolova 2014; Ding, Kulkarni, and Yekhanin 2017). We evaluate estimator performance across three representative data distributions, characterized by Zipf concentration parameters: $s = 0.01$ **(near-uniform)**, $s = 1.3$ **(moderately skewed)**, and $s = 2.5$ **(highly skewed)**. Figure 2 complements this view by fixing the number of users to $N = 10^6$ and instead varying the privacy budget ϵ, providing insights into estimator behavior across different privacy regimes.



(a) Mean Squared Error (MSE), $K = 10000, \epsilon = 4$



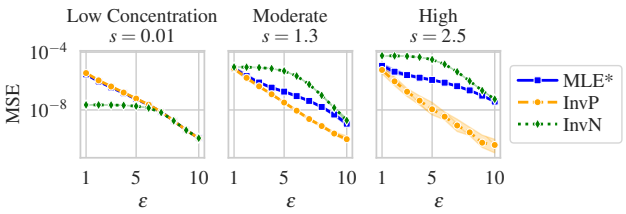
(b) Negative Log-Likelihood, $K = 10000, \epsilon = 4$

Figure 1: Performance of MLE$\star$, InvP, and InvN across different data distributions ($s \in \{0.01, 1.3, 5.0\}$) for fixed $K = 10{,}000$ and $\epsilon = 4.0$.

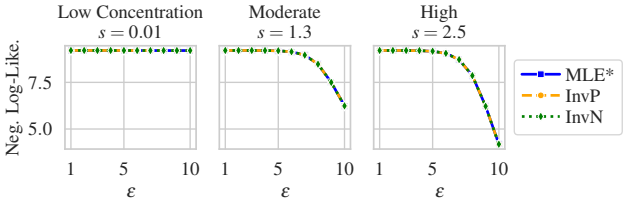
Discussion. For the MSE metric, these results highlight the complexity of estimator behavior across different privacy and data regimes. As shown in Figures 1a and 2a, the relative performance of InvP and InvN varies significantly with both the distributional shape (controlled by s) and system parameters (N, ϵ).

We observe that in some settings, one estimator clearly dominates: for instance, InvP consistently performs best under high concentration ($s = 2.5$), while InvN is clearly superior in the near-uniform regime ($s = 0.01$). In other cases, the ranking between InvP and InvN can change depending on ϵ or N, with one starting off stronger but eventually being overtaken as conditions change.

Amidst this variability, MLE$\star$ demonstrates consistently strong and stable performance. It is consistently "sandwiched" between the two baselines, i.e., never the worst, often close to the best. This behavior underscores the **robustness** of MLE$\star$: it adapts well across a wide range of scenarios without requiring prior knowledge of the underlying



(a) Mean Squared Error (MSE), $K = 10000$, $N = 10^6$



(b) Negative Log-Likelihood, $K = 10000$, $N = 10^6$

Figure 2: Performance of MLE$\star$, InvP, and InvN across different data distributions ($s \in \{0.01, 1.3, 5.0\}$) for fixed $K = 10{,}000$ and $N = 10^6$.

data distribution or the optimal estimator for a given configuration. As such, MLE$\star$ is a reliable default choice when performance must be maintained under uncertainty.

Beyond MSE, as shown in both Figures 1b and 2b, MLE$\star$ consistently achieves the lowest Negative Log-Likelihood across all configurations, regardless of the data distribution, sample size, or privacy budget. This is expected, as our estimator is explicitly derived via maximum likelihood and optimized to minimize this very objective. In contrast, InvP and InvN, which are not likelihood-based, often result in poorer fit to the observed data in terms of Negative Log-Likelihood, even when they perform well under MSE. These results confirm that MLE$\star$ not only offers robust accuracy but is also statistically well-calibrated to the data generation process.

MLE$\star$ as a Robust Default Estimator

Real-world distributions are unknown and often highly variable. Across our extensive evaluations, no single baseline estimator consistently dominates: InvP and InvN alternate in performance depending on the privacy budget, sample size, and distribution skew. In contrast, our likelihood-based estimator, MLE$\star$, remains reliably close to the best in all scenarios. This consistency makes our MLE$\star$ **a safe and robust default** for practical deployment.

6.3 Experiments on Real-World Data

To assess the applicability of our estimators in practical settings, we evaluate their performance on two real-world datasets: KOSARAK[1] and ACSINCOME (Ding et al. 2021). In both cases, we fix the number of users N to the dataset size and vary the privacy parameter in the range $\epsilon \in \{1, \ldots, 10\}$. We set the domain size K based on the number of unique values in the selected column, as described below.

Kosarak. This dataset consists of clickstream data from a Hungarian online news website. Each record represents a user and the set of URLs they clicked. We extract a single histogram (1st reported URL per user) over all 41,270 unique URLs, resulting in a domain size of $K = 41{,}270$.

ACSIncome. This dataset is derived from the US Census. We extract histograms based on two distinct attributes:

- **PUMA:** The *Public Use Microdata Area code*, a geographic identifier with codes ranging from 100 to 70,301. We thus set the domain size to $K = 70{,}201$.
- **PINCP:** The *Total Person's Income*, is a continuous variable ranging from 100 to 1,423,000. We thus set the domain size to $K = 1{,}423{,}000$.

Evaluation. For each dataset and attribute, we evaluate the MSE as a function of the privacy parameter ϵ. We also include the true underlying histogram for visualization. These experiments highlight the behavior of the estimators in high-dimensional, real-world scenarios, where the distributions can be highly skewed or sparse.

Figure 3 summarizes the results. Each subfigure displays: (left) the true data distribution and (right) the MSE as a function of the privacy budget ϵ, for the KOSARAK dataset (a), and for the ACSINCOME dataset using the PUMA (b) and PINCP (c) attributes, respectively.

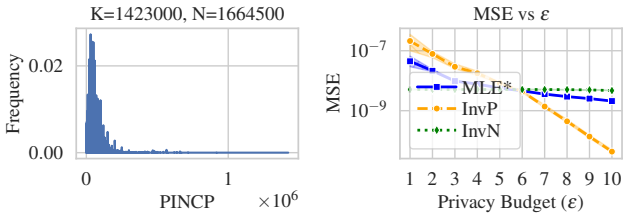
Discussion. The findings from real-world datasets further reinforce the insights observed in our controlled synthetic experiments (Section 6.2). As shown in Figure 3, the ACSINCOME distributions (PUMA and PINCP) exhibit moderate concentration, similar to the synthetic setting with Zipf parameter $s = 1.3$. In contrast, the KOSARAK dataset shows a highly peaked and sparse distribution, closely resembling the highly concentrated regime with $s \geq 2.5$.

In these respective settings, InvP performs best under strong concentration (as in KOSARAK), while InvP and InvN alternate in the moderately skewed regime (ACSINCOME). It is important to note, however, that these observations are based on fixed domain size K and a fixed number of users N. Despite such fluctuations in relative performance, MLE$\star$ consistently remains close to the best-performing estimator across all cases. Its stable behavior across datasets, metrics, and privacy levels reinforces our central conclusion: MLE$\star$ is a robust and dependable choice, particularly in practical settings where the true data distribution is unknown and must be inferred from obfuscated data using the RR mechanism.

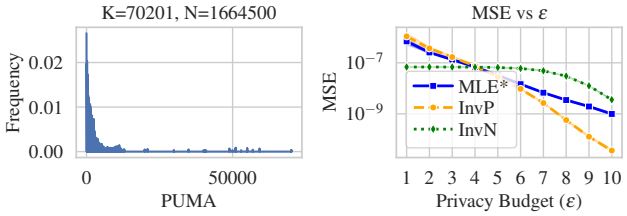
6.4 The exact MLE vs. IBU's approximation

In the following, we experimentally compare our MLE$\star$, which gives the exact MLE for the original distribution, and the existing IBU method, which iteratively approximates this estimate. This comparison verifies the correctness of MLE$\star$ empirically, and also shows the computational savings that it achieves. Precisely, we measure the squared error between IBU's estimate at each iteration and the exact

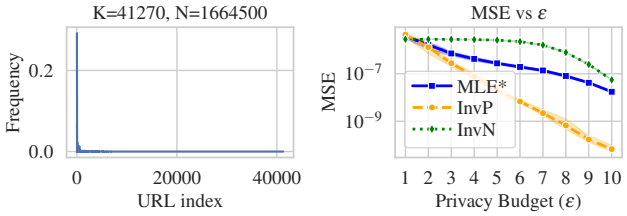
[1] http://fimi.uantwerpen.be/data/



(a) ACSIncome (PINCP)



(b) ACSIncome (PUMA)



(c) Kosarak

Figure 3: Performance of `MLE⋆`, `InvP`, and `InvN` across three different real-world distributions.

result of `MLE⋆`. Figure 4, shows the plot of this error in four experiments performed with two original Zipf distributions ($s = 0.01, 1.3$), two sample sizes ($n = 10^4, n = 50 \times 10^4$), and different privacy levels ($\epsilon = 0.5, 1.0, 2.0$). In each experiment, we sample n data points from the original distribution, sanitize these samples using an RR mechanism (with $K = 500$), and finally run `IBU` to approximate the MLE of the original distribution.

It can be seen that the error of `IBU` converges to 0 with more iterations, hence confirming the correctness of `MLE⋆` empirically. The speed of this convergence depends on the setting as follows. For $n = 10^4$, and a moderate level of privacy, e.g. $\epsilon = 1.0$, `IBU` requires about 7000 iterations to approach the exact value obtained via `MLE⋆`, while with a stronger level of privacy ($\epsilon = 0.5$), it requires a larger number of iterations (around 25000) to obtain the same approximation level. For the larger population ($n = 50 \times 10^4$), many more iterations are needed for `IBU` to reach a reliable approximation. In particular, the moderate privacy level $\epsilon = 1.0$ requires about 20000 iterations, while for the stronger privacy ($\epsilon = 0.5$), even 40000 iterations are not enough to reach the same approximation precision. This indeed shows the computational superiority of `MLE⋆`.

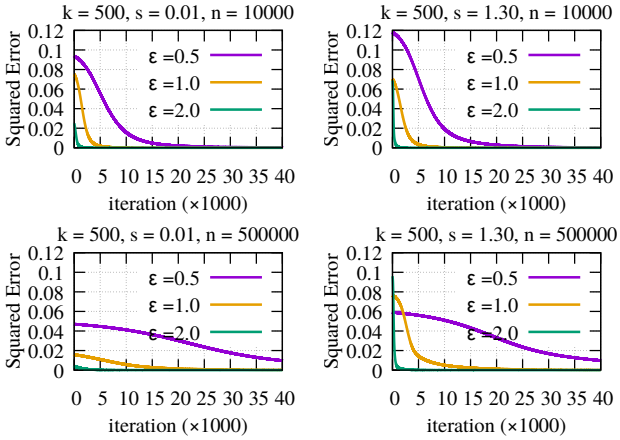


Figure 4: The squared error of the `IBU`'s estimates relative to `MLE⋆`, with the number of iterations.

7 Conclusion

We addressed several limitations in frequency estimation under randomized response.

First, we derived a formula for the Maximum Likelihood Estimator and proved its correctness both theoretically and empirically. We also proved the uniqueness of the MLE. Our derivation corroborates the results of (Kairouz, Bonawitz, and Ramage 2016), Supplementary Material, Section F, and, since the formula does not involve limits of any kind, it can be used to better understand the analytical properties of the MLE. Second, we proposed an algorithm, `MLE⋆`, that computes the MLE in $O(K \log K)$ using the formula we found, vastly outperforming the iterative procedure IBU in speed while also providing an exact output as opposed to an approximation. Third, we conducted extensive empirical comparisons of the most prominent valid estimators (`InvN`, `InvP` and the MLE using `MLE⋆`) to shed light on their behavior and trade-offs under different scenarios. We found that in terms of mean squared error, the projection method `InvP` outperforms `InvN` on average when the unknown target distribution is highly concentrated and vice-versa. Furthermore, while `InvP` and `InvN` outperform each other under specific conditions, `MLE⋆` consistently stays in between, which makes it robust in that it is never the worst of the three. Its validity, robustness, and efficiency make `MLE⋆` a reliable default choice when the data distribution is unknown and performance under uncertainty matters.

Since many practical mechanisms are built upon or inspired by RR, an interesting direction for future research is to extend our analysis of the formula of `MLE⋆` to other LDP frequency estimation protocols (Wang et al. 2017), in particular to *longitudinal LDP protocols* (Erlingsson, Pihur, and Korolova 2014; Arcolezi et al. 2022; Ding, Kulkarni, and Yekhanin 2017), which protect users' data across repeated collections and are highly relevant in real-world telemetry and monitoring applications. Another future direction is to combine different estimators to enrich the analysis. Finally, a bias-variance is included in the code repository (Pinzón et al. 2025).

References

Agrawal, D.; and Aggarwal, C. C. 2001. On the design and quantification of privacy preserving data mining algorithms. In *Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, 247–255.

Agrawal, R.; Srikant, R.; and Thomas, D. 2005. Privacy preserving OLAP. In *Proceedings of the 2005 ACM SIGMOD international conference on Management of data*, 251–262.

Arcolezi, H. H.; Pinzón, C.; Palamidessi, C.; and Gambs, S. 2022. Frequency estimation of evolving data under local differential privacy. *arXiv preprint arXiv:2210.00262*.

Berger, J. O. 1990. On the inadmissibility of unbiased estimators. *Statistics & probability letters*, 9(5): 381–384.

Cheu, A.; Smith, A.; Ullman, J.; Zeber, D.; and Zhilyaev, M. 2019. Distributed differential privacy via shuffling. In *Annual international conference on the theory and applications of cryptographic techniques*, 375–403. Springer.

Dempster, A. P.; Laird, N. M.; and Rubin, D. B. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society: series B (methodological)*, 39(1): 1–22.

Ding, B.; Kulkarni, J.; and Yekhanin, S. 2017. Collecting telemetry data privately. *Advances in Neural Information Processing Systems*, 30.

Ding, F.; Hardt, M.; Miller, J.; and Schmidt, L. 2021. Retiring adult: New datasets for fair machine learning. *Advances in neural information processing systems*, 34: 6478–6490.

Duchi, J. C.; Jordan, M. I.; and Wainwright, M. J. 2013. Local privacy and statistical minimax rates. In *2013 IEEE 54th annual symposium on foundations of computer science*, 429–438. IEEE.

ElSalamouny, E.; and Palamidessi, C. 2020. Generalized iterative bayesian update and applications to mechanisms for privacy protection. In *2020 IEEE European Symposium on Security and Privacy (EuroS&P)*, 490–507. IEEE.

ElSalamouny, E.; and Palamidessi, C. 2025. On the Consistency and Performance of the Iterative Bayesian Update. arXiv:2508.09980.

Erlingsson, Ú.; Feldman, V.; Mironov, I.; Raghunathan, A.; Talwar, K.; and Thakurta, A. 2019. Amplification by shuffling: From local to central differential privacy via anonymity. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, 2468–2479. SIAM.

Erlingsson, Ú.; Pihur, V.; and Korolova, A. 2014. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, 1054–1067.

Hay, M.; Rastogi, V.; Miklau, G.; and Suciu, D. 2009. Boosting the accuracy of differentially-private histograms through consistency. *arXiv preprint arXiv:0904.0942*.

Jacobs, B. 2021. Learning from What's Right and Learning from What's Wrong. *arXiv preprint arXiv:2112.14045*.

Kairouz, P.; Bonawitz, K.; and Ramage, D. 2016. Discrete distribution estimation under local privacy. In *International Conference on Machine Learning*, 2436–2444. PMLR.

Lee, J.; Wang, Y.; and Kifer, D. 2015. Maximum likelihood postprocessing for differential privacy under consistency constraints. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 635–644.

Pinzón, C.; H. Arcolezi, H.; ElSalamouny, E.; and Massot, L. 2025. Code Repository for the paper Estimating the True Distribution of Data Collected with Randomized Response. https://github.com/caph1993/mle-of-randomized-response. Accessed: 2025-11-15.

Pinzón, C.; and Palamidessi, C. 2025. Jeffrey's update rule as a minimizer of Kullback-Leibler divergence. arXiv:2502.15504.

Wang, T.; Blocki, J.; Li, N.; and Jha, S. 2017. Locally differentially private protocols for frequency estimation. In *26th USENIX Security Symposium (USENIX Security 17)*, 729–745.

Warner, S. L. 1965. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American statistical association*, 60(309): 63–69.

Ye, Y.; Wang, T.; Zhang, M.; and Feng, D. 2025. Revisiting EM-based Estimation for Locally Differentially Private Protocols. In *32nd Annual Network and Distributed System Security Symposium, NDSS*, 24–28.