Marginal Release Under Local Differential Privacy

Tejas Kulkarni (University Of Warwick), Graham Cormode (University Of Warwick), Divesh Srivastava (AT&T Labs-Research)

Introduction

Marginal statistics of multi-dimensional data are the workhorse of data analysis.

Applications of marginals range from finding correlations in the data to fitting sophisticated prediction models.

We provide a set of algorithms for materializing marginal statistics under the strong model of local differential privacy.

We prove theoretical bounds on the accuracy of marginals, and perform empirical evaluation for tasks such as modeling and correlation testing.

Background

Local Differential Privacy (LDP) [4] requires that the output of *every* user meets the ε-differential privacy guarantee.

This is typically achieved by perturbing the user's output with some probability, e.g. flipping a bit.

LDP for data analysis has been deployed at large scale in systems from Google [5], Apple [2] and Microsoft [3].

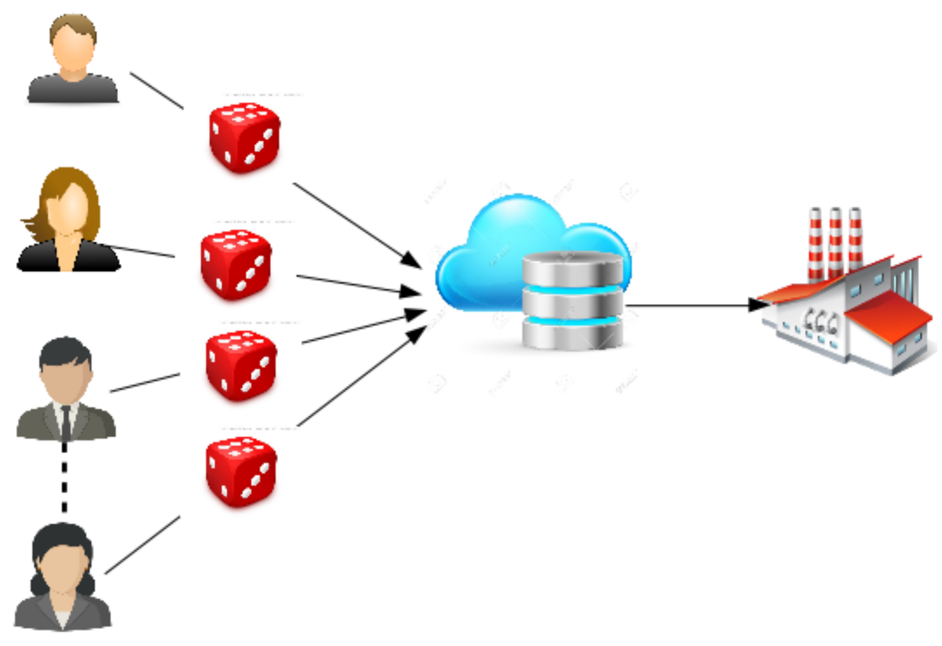

Figure 1: Untrusted Aggregation

Problem Statement

Suppose a research organization wants to collect sensitive data from each user through an online survey/smartphone app.

To ensure total privacy, the organization won't collect raw data, but will define a protocol for users to mask their own data.

Each user is asked to answer d **sensitive binary** questions, e.g. gender, overweight, smoker, high BP.

The **aggregator's goal** is to find associations (a.k.a marginals) between arbitrary subsets of k questions/attributes (out of the total $\binom{d}{k}$ subsets).

Overweight/High BP	Y	N
Y	0.55	0.15
N	0.10	0.20

Figure 2: A 2-way marginal

Our Contributions

We develop 6 perturbation algorithms combining views of the data (global/local view) and basis transforms (identity and Hadamard transform).

We show the mathematical relationship between various parameters (dimension d, marginal size k, privacy parameter ε, population size N) and the accuracy of aggregation. The error in aggregation is $\propto \frac{1}{\varepsilon}\sqrt{\frac{1}{N}}$.

Algorithm Design

Our methods differ depending on how they view the data:

- **Input Based (Inp*).** The aggregator gathers information from each user to reconstruct the full distribution, then projects down to the marginal of interest at query time.
- **Marginal Based (Marg*).** Each user evaluates a random marginal, and releases (perturbed) data on it, so each marginal is reconstructed independently.

Applying Data Transformations. Each user's input can be represented as a vertex in a d-dimensional Hamming cube:

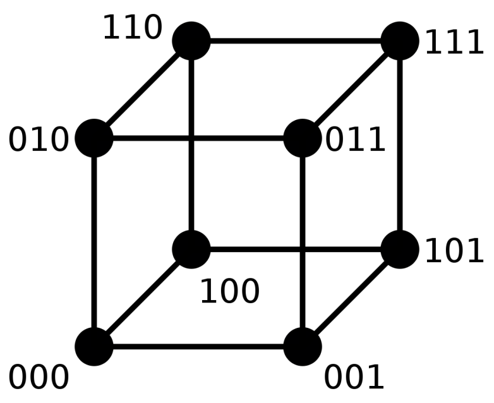

Figure 3: Hamming Cube

Inputs are sparse, so we consider transforming the input to make it more dense.

We can apply the *Hadamard Transform* (HT), a discrete Fourier transform for the d-dimensional Hamming cube.

HT is **linear**, so Hadamard coefficients for the whole population are formed as the sum of coefficients from each individual.

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{pmatrix}$$

Figure 4: Hadamard Transformation Matrix

Our best algorithm (InpHT) has each user report *one* randomly sampled and perturbed Hadamard coefficient of their input.

The aggregator builds an unbiased estimate of the HT of the input, and uses this to reconstruct any required marginal.

Theorem [1]: Only $\mathbb{C} = \sum_{i \leq k} \binom{d}{i} \leq d^k$ Hadamard coefficients are needed to evaluate *any* of the marginals involving $\leq k$ attributes.

Evaluation

Data set. We take the 2013 NYC Taxi Data set, and create 8 binary attributes:

Attribute	Explanation
CC	Has customer paid using credit card?
Toll	Has customer paid toll?
Far	Is journey distance ≥ 10 miles?
Night_pick	Is pickup time ≥ 8 PM?
Night_drop	Is drop off time ≤3 AM?
M_pick	Is trip origin within Manhattan?
M_drop	Is trip destination within Manhattan?
Tip	Is tip paid ≥ 25% of the total fare?

Figure 5: Binary attributes of NYC taxi data

We apply our methods to statistical/machine learning tasks like χ^2 **test of independence** and **Bayesian Modeling** (approximating a high dimensional marginal using low dimensional ones).

Marginal reconstruction. We find that the Hadamard-based method on the full input (InpHT) gives the best accuracy in reconstructing marginals. The second best method materializes marginals directly (MargPS).

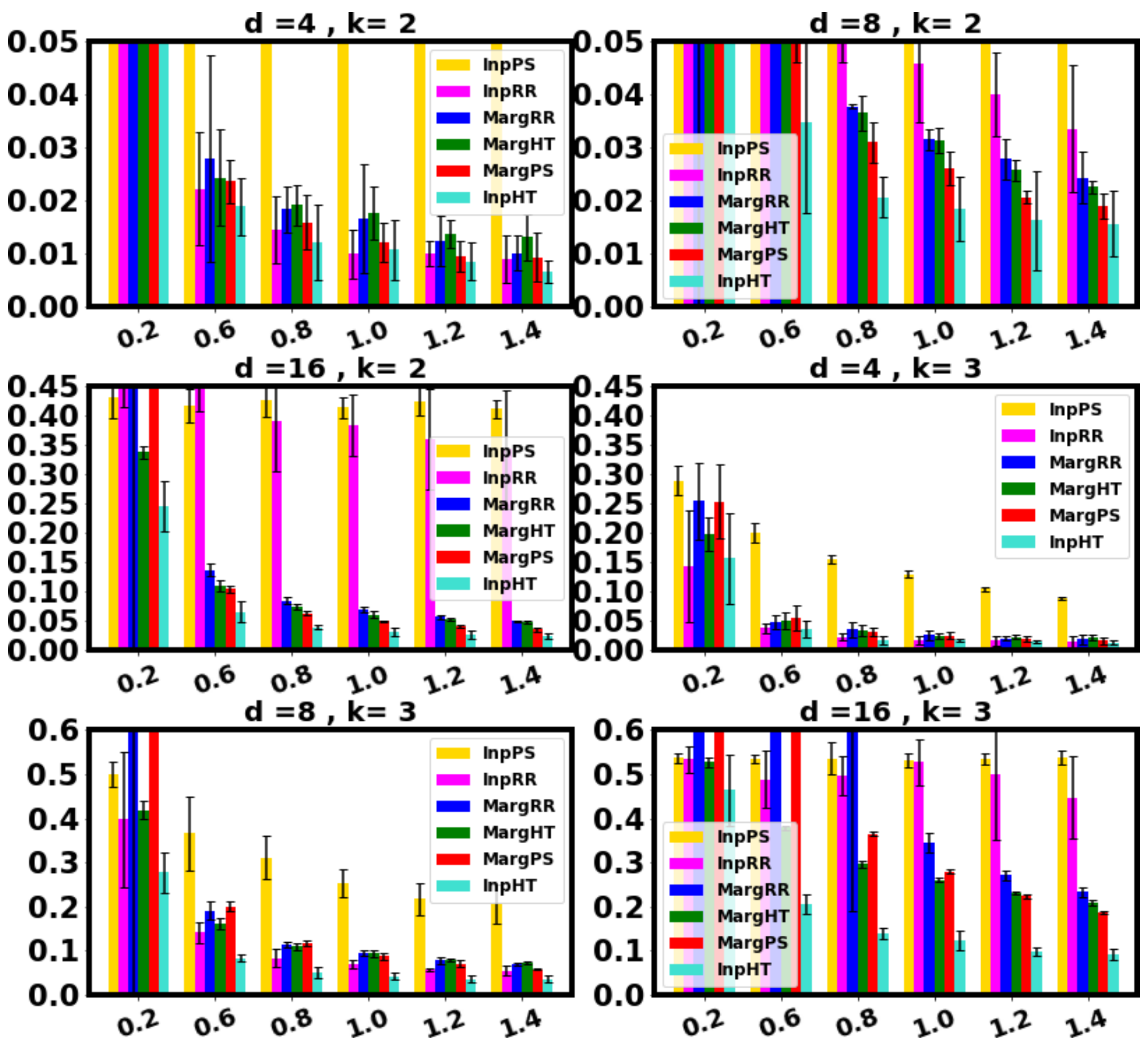

Figure 6: Mean total variation between true and reconstructed marginals as ε varies

χ^2-test of independence using 2-way marginals. We use the materialized marginals to run a χ^2 test for significance of correlation. We observe that InpHT obtains test values closer to the correct ones than the alternative method MargPS.

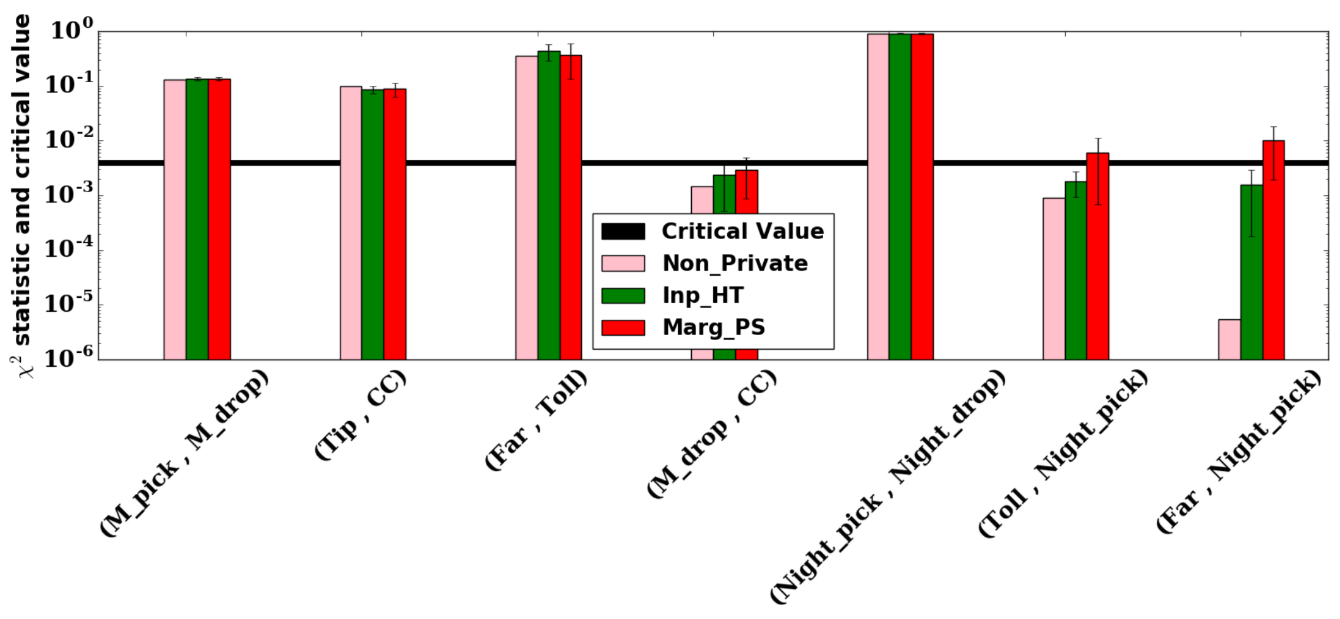

Figure 7: χ^2 test statistics for $N = 256K$ taxi trips with $\varepsilon = 1.1$

Concluding Remarks

We show that accurate marginal reconstruction is possible under the local model of differential privacy. Open problems include:

- **Extend to non-binary data.** Encoding categoric variables as binary is a first step.
- **More complex data analytics.** It is open to more directly build rich models accurately under LDP.

Acknowledgements. This work is supported in part by a grant from AT&T, European Research Council grant ERC-2014-CoG 647557, Marie Curie CIG 618202, and The Alan Turing Institute under the EPSRC grant EP/N510129/1.

References

[1] B. Barak, K. Chaudhuri, C. Dwork, S. Kale, F. McSherry, and K. Talwar. Privacy, accuracy, and consistency too: a holistic solution to contingency table release. In *PODS*, pages 273–282. ACM, 2007.

[2] Differential Privacy Team, Apple. Learning with privacy at scale. 2017.

[3] B. Ding, J. Kulkarni, and S. Yekhanin. Collecting telemetry data privately. In *NIPS*, pages 3574–3583, 2017.

[4] J. Duchi, M. Jordan, and M. Wainwright. Local privacy and statistical minimax rates. In *FOCS*, 2013.

[5] U. Erlingsson, V. Pihur, and A. Korolova. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *ACM CCS*, 2014.