Evaluation of Label DP (Randomized Response) Mechanisms on Binary Conversion Models

Badih Ghazi (Algorithmic Privacy Team, Google Research)

Differential Privacy (DP)

[Dwork et al.'06]

Algorithm A is ε-DP if for all output values o, and two adjacent datasets D, D':

$\Pr[A(D) = o] \leq e^{\varepsilon} \cdot \Pr[A(D') = o]$

Adjacency relation corresponds to privacy unit.

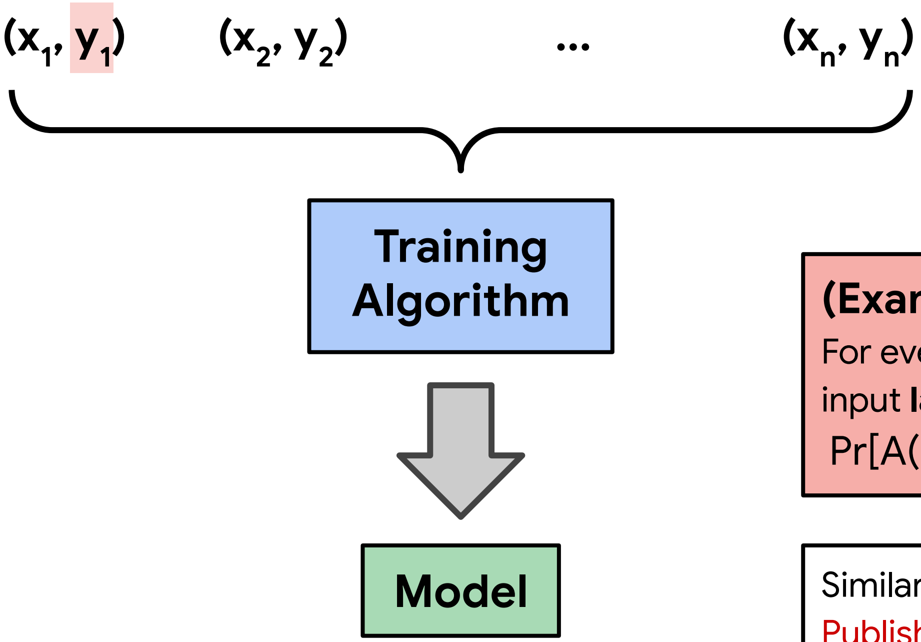
Possible privacy units for conversion measurement

- Impression x Time
- User x Publisher x Time
- User x Advertiser x Time
- User x Time

Note: In our experiments, the privacy budget is for a single ad-tech (a.k.a. report collector), so "x Ad-Tech" is implicit in the privacy unit.

Training with *Label* Differential Privacy

sensitive/unknown data



(x_1, y_1) (x_2, y_2) … (x_n, y_n)

Training Algorithm

Model

(Example-level) ε-Label-DP

For every datasets D, D’ that differ by a single input **label** and every output model o,

$$\Pr[A(D) = o] \leq e^{\varepsilon} \cdot \Pr[A(D') = o]$$

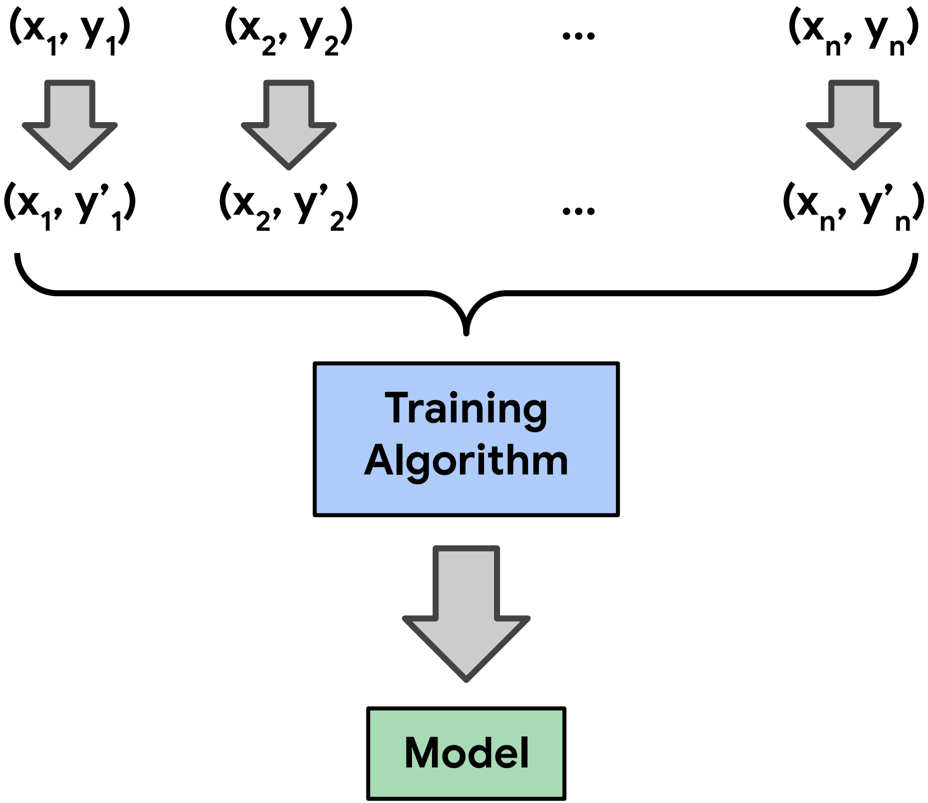
Similarly, for Impression x Time, User x Publisher x Time, User x Advertiser x Time, and User x Time privacy units

Binary Randomized Response

RR_p *[Warner'65]*

Given $\mathbf{y_i}$ in {0, 1}:
$\mathbf{y'_i = y_i}$ with probability **1 - p**
Otherwise, $\mathbf{y'_i}$ is a random label in {0, 1}.

RR_p is ε-DP for $p = 2/(e^{\varepsilon}+1)$.



(x_1, y_1) (x_2, y_2) ... (x_n, y_n)

(x_1, y'_1) (x_2, y'_2) ... (x_n, y'_n)

Training Algorithm

Model

If y'_i is obtained by applying RR_p independently to y_i, then output model is ε-Label-DP.

Conversion Modeling

Prediction task:

Given an impression with associated ad, contextual, and publisher 1P user features, predict whether the impression will get an attributed conversion.
(Prediction used as input in automated bidding when deciding how much to bid on ad in auction.)

Metrics:

- AUC (Area under the Curve):
  - Plot *True Positive Rate* against *False Positive Rate*, and compute area under it.
  - AUC = probability that the classifier ranks a randomly chosen positive example higher than a randomly chosen negative one.
  - Random guessing: AUC = 0.5
  - Perfect prediction utility: AUC = 1
- AUC-Loss = 1 - AUC
- Relative Change in AUC-Loss (in %) = (AUC-Loss - AUC-Loss$_{\text{baseline}}$) / AUC-Loss$_{\text{baseline}}$ * 100
  - baseline is without DP

Handling Multiple Impressions Per Privacy Unit

Example: Consider User x Time privacy unit.

Cap number of impressions per user and time period to K (keeping K random impressions, or the K first impressions). Then, we have multiple options including:

1. For each user, set the privacy budget per impression to ε/K.
2. For each user i with $K_i \leq K$ impressions, set the privacy budget per impression to $ε/K_i$.

Both options satisfy ε-Label-DP for User x Time privacy unit.

Similar options hold for User x Advertiser x Time and User x Publisher x Time privacy units.

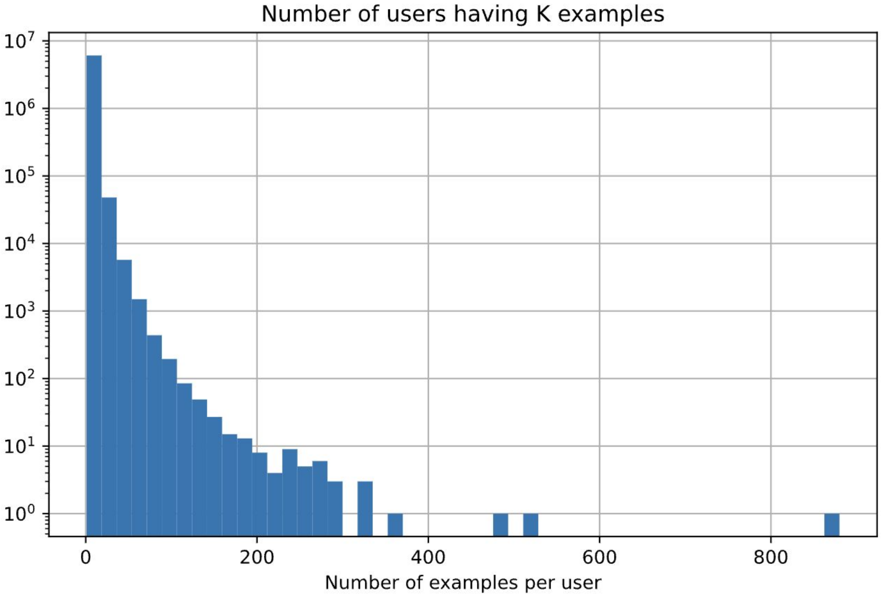
For Impression x Time privacy unit, no capping is needed. RR is applied privacy budget ε.

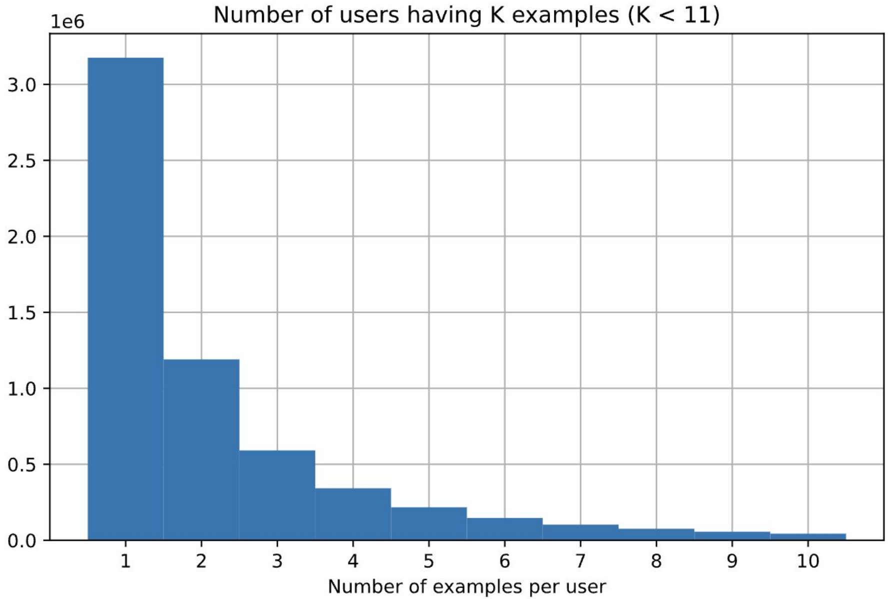
Criteo Attribution Modeling for Bidding Dataset

https://ailab.criteo.com/criteo-attribution-modeling-bidding-dataset/

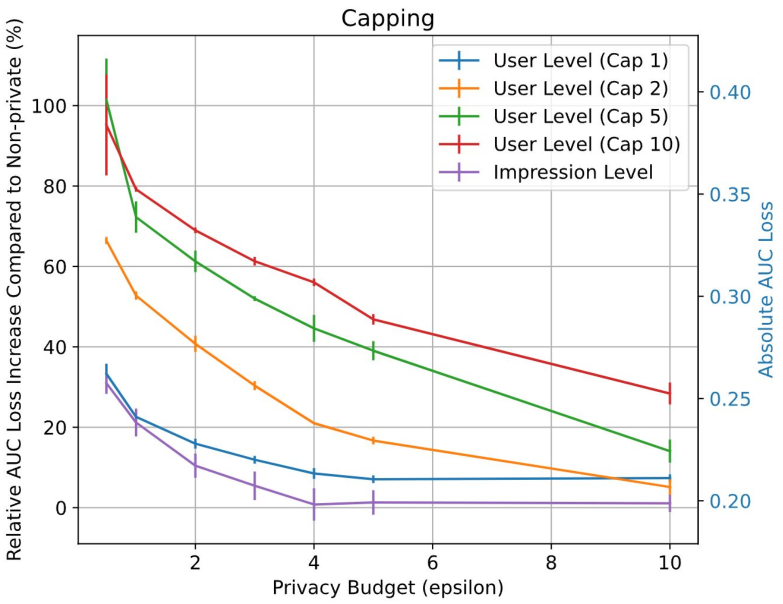
- Sample of 30 days of Criteo live traffic data.
- Each example corresponds to a click and contains:
  - **Features:** campaign ID, 9 contextual features, and the cost paid for the display.
  - **Label:** a 0/1 field indicating whether there was a conversion in the 30 days after the click and that is last-touch attributed to this click.
  - **User ID:** can be used to evaluate User x Time privacy unit.
- Number of rows is 5,947,563. Conversion rate (under last-touch attribution) is 6.74%.
- Feed-forward neural network
  - Embedding dimension is 8
  - Hidden dimensions are 128 x 64.
- Tune hyperparameters with optimizer in {rmsprop, adam, sgd}, learning rate in {0.0005, 0.0008, 0.001, 0.002, 0.005}, and training epochs in {100, 200}.

Criteo Attribution Modeling for Bidding – Statistics





Criteo Attribution Modeling for Bidding – Evaluation Results



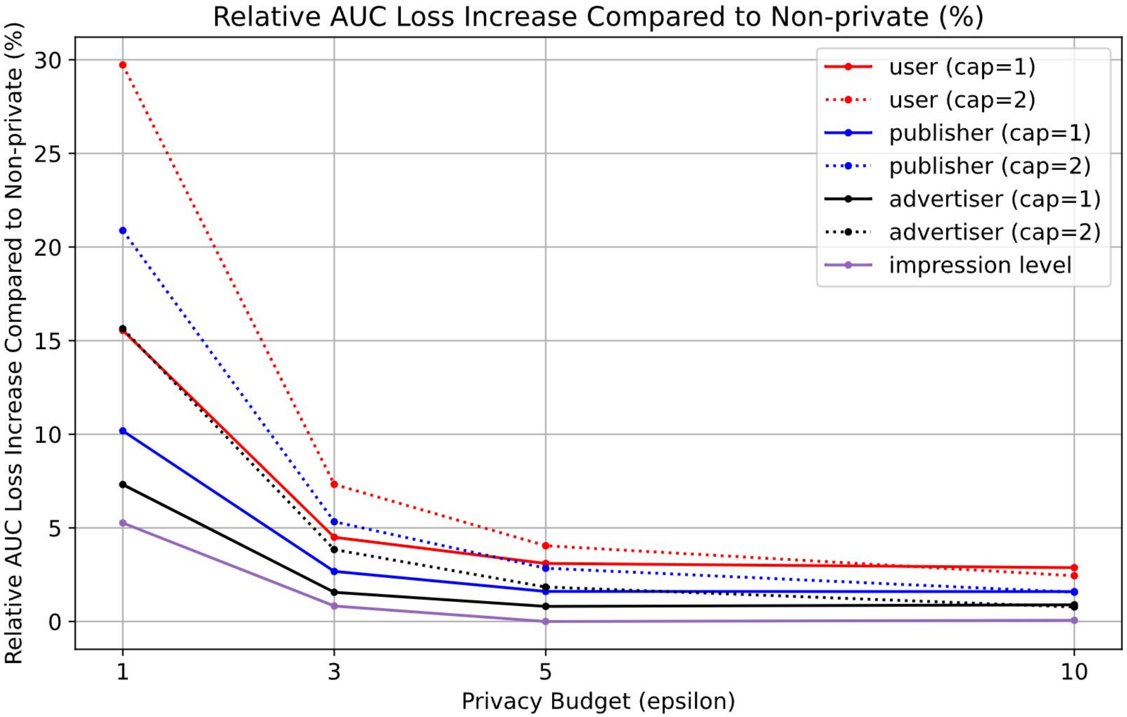
Notes

- For Impression x Time privacy unit and ε = 4, relative AUC loss is 0.79%.
- For User x Time privacy unit with ε = 4, smallest relative AUC loss is 8.51%.
- For User x Time privacy unit, smaller loss is achieved by increasing caps as we increase ε.

Proprietary Ads Dataset

- App install ads
- Contains data from multiple advertisers and publishers.
  - Can be used to evaluate User x Time, User x Publisher x Time, and User x Advertiser x Time privacy units
- **Features:** use categorical features, and pass the concatenation of their embeddings through multiple layers of a fully connected feedforward neural network.
- **Labels:** 0/1 corresponding to installs (= conversions)

Proprietary Ads Dataset – Evaluation Results



Notes

- For Impression x Time privacy unit and ε = 3, relative AUC loss is 0.83%.
- For User x Time privacy unit and ε = 3, smallest relative AUC loss is 4.50%.
- For User x Publisher x Time privacy unit and ε = 3, smallest Relative AUC loss is 2.67%.
- For User x Advertiser x Time unit with ε = 3, best Relative AUC loss is 1.56%.
- In this experiment, for same ε,
  loss for Impression x Time
  < loss for User x Advertiser x Time
  < loss for User x Publisher x Time
  < loss for User x Time.

Limitations and Future Directions

- Utility might be improvable for binary conversion models
  - Debiased loss functions
- Evaluation focused on binary conversion models
  - Label DP algorithms (and RR in particular) can be extended to non-binary predictions (such as predicting number of conversions, and conversion value etc.)
- Evaluation was done offline and assumed one reporting window
  - Online performance can be impacted by delays, which might require multiple reporting windows (and more noise for the same privacy)

References

- Calibrating noise to sensitivity in private data analysis
  Dwork, McSherry, Nissim, Smith TCC 2006
- Randomized response: A survey technique for eliminating evasive answer bias
  Warner, JASA 1965
- Deep learning with label differential privacy
  Ghazi, Golowich, Kumar, Manurangsi, Zhang
  NeurIPS 2021
- Antipodes of label differential privacy:PATE and ALIBI
  Malek, Mironov, Prasad, Shilov, Tramèr
  NeurIPS 2021
- Regression with Label Differential Privacy
  Ghazi, Kamath, Kumar, Leeman, Manurangsi, Varadarajan, Zhang
  ICLR 2023