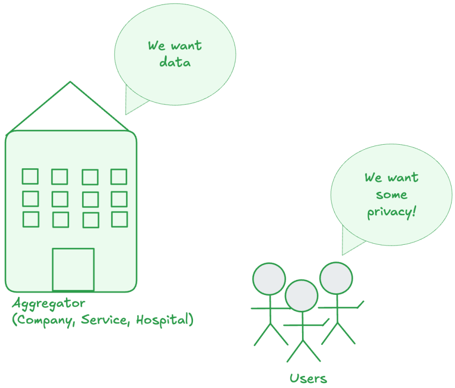
Privacy-Preserving Conformal Prediction Under Local Differential Privacy

Coby Penso   Bar Mahpud   Jacob Goldberger   Or Sheffet

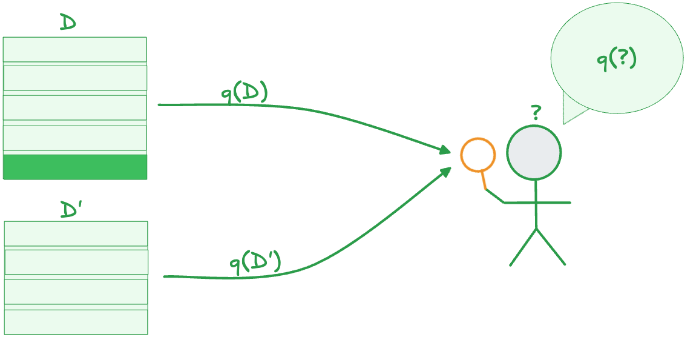
Faculty of Engineering, Bar-Ilan University

Conformal And Probabilistic Prediction With Applications
September 2025

Differential Privacy – Motivation



Differential Privacy – Motivation



Local Differential Privacy

ε-local differential privacy [DR13]

A discrete randomized mechanism $A(\cdot)$ is ε-LDP if for any pair of input labels $y, y' \in \mathcal{Y}$ and any output z,

$$\Pr[A(y) = z] \;\leq\; e^{\varepsilon} \Pr[A(y') = z].$$

This means that any two possible labels are (roughly) indistinguishable from the aggregator's perspective.

k-RR (k-ary randomized response)

k-RR [War65]

For a label $y \in \{1, \ldots, k\}$, it outputs a noisy version $\tilde{y}$:

$$p(\tilde{y} \mid y) = 1_{\{\tilde{y}=y\}}(1-\beta) + \frac{\beta}{k}, \quad \beta = \frac{k}{k-1+e^{\epsilon}}$$

k-RR satisfies ϵ-local-differential privacy.

Post-processing property of differential privacy [Dwo06]

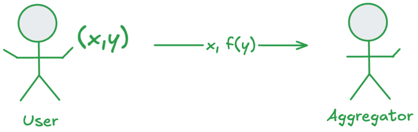
Let $M : \mathcal{D} \to \mathcal{O}$ be an ε-differentially private mechanism. For any function f that does not depend on the input dataset D, define $M'(D) = f(M(D))$. Then M' is also ε-differentially private.

CP under Local Differential Privacy

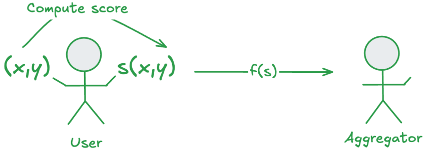
In local differential privacy, the aggregator is untrusted.

What we protect?

- Protecting only labels y, revealing input x allowed
  $\rightarrow$ Local differential privacy on labels

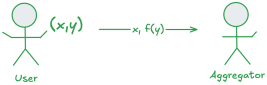


- Protecting labels and inputs x, y
  $\rightarrow$ Local differential privacy on scores



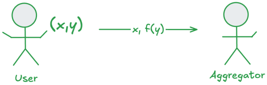
CP under Local Differential Privacy - Labels

- Goal: Conformal prediction procedure that satisfies local differential privacy on labels.
- Two challenges:
  - How to maintain label privacy?
  - Label privacy comes with a cost. How to output a correct conformal prediction result given label privacy?



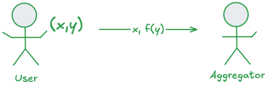
CP under Local Differential Privacy - Labels

- Goal: Conformal prediction procedure that satisfies local differential privacy on labels.
- Two challenges:
  - How to maintain label privacy?
    randomize labels at source using k-RR mechanism.
  - Label privacy comes with a cost. How to output a correct conformal prediction result given label privacy?



CP under Local Differential Privacy - Labels

- Goal: Conformal prediction procedure that satisfies local differential privacy on labels.
- Two challenges:
  - How to maintain label privacy?
    randomize labels at source using k-RR mechanism.
  - Label privacy comes with a cost. How to output a correct conformal prediction result given label privacy?
    conformal prediction with noisy labels (e.g. NACP [PGF25]).



CP under Local Differential Privacy - Labels

Altogether,

1. Users apply k-RR
2. Aggregator runs Noise-Aware CP

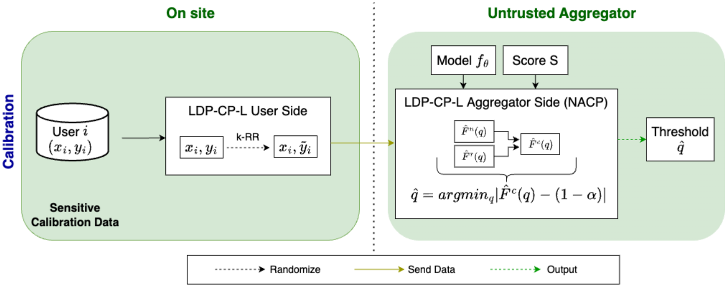
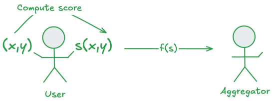


Figure: Local Differential Private Conformal Prediction on Labels (LDP-CP-L).

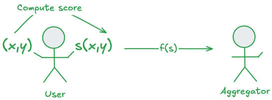
CP under Local Differential Privacy - Scores

- Goal: Conformal prediction procedure that satisfies local differential privacy on scores.
- Challenges:
  - How to maintain score privacy?
  - Scores are continues and not categorical, k-RR not appliciable. Scores are sensitive, noisy scores are not applicable
  - How to output a correct conformal prediction result given score privacy?



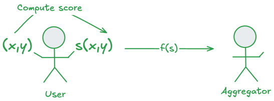
CP under Local Differential Privacy - Scores

- Goal: Conformal prediction procedure that satisfies local differential privacy on scores.
- Challenges:
  - How to maintain score privacy?
    randomize scores or response at source.
  - Scores are continues and not categorical, k-RR not applicable?
    Scores are sensitive, noisy scores are not applicable
  - How to output a correct conformal prediction result given score privacy?



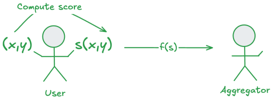
CP under Local Differential Privacy - Scores

- Goal: Conformal prediction procedure that satisfies local differential privacy on scores.
- Challenges:
  - How to maintain score privacy?
    randomize scores or response at source.
  - Scores are continues and not categorical, k-RR not appliciable?
    Scores are sensitive, noisy scores are not applicable
    randomize the response to the aggregator's query instead of scores.
  - How to output a correct conformal prediction result given score privacy?



CP under Local Differential Privacy - Scores

- Goal: Conformal prediction procedure that satisfies local differential privacy on scores.
- Challenges:
  - How to maintain score privacy?
    randomize scores or response at source.
  - Scores are continues and not categorical, k-RR not appliciable? Scores are sensitive, noisy scores are not applicable randomize the response to the aggregator's query instead of scores.
  - How to output a correct conformal prediction result given score privacy?
    use $(1-\alpha)$-quantile local differential private algorithm.



CP under Local Differential Privacy - Scores

Altogether,

1. Aggregator iteratively queries sub-groups of users until finds the $(1-\alpha)$-quantile
2. Users apply 2-RR on Aggregator query $1(s < q^j)$

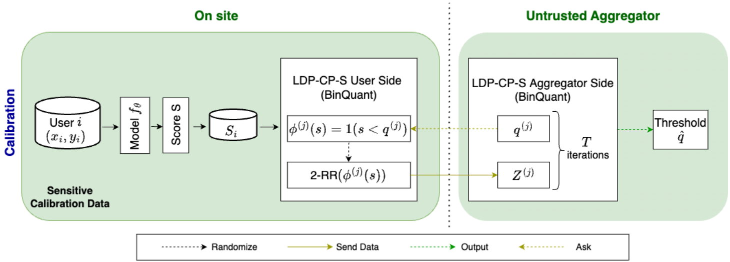


Figure: Local Differential Private Conformal Prediction on Scores (LDP-CP-S).

Theorems

Theorem - LDP-CP-L

Local Differentially Private Conformal Prediction under Labels Privacy. Fix $\alpha, \delta, \Delta > 0$. There exists an ϵ-local differentially private algorithm that draws $n = O\left(\frac{\log(1/\delta)}{\Delta^2 h^2}\right) \sim D$, where $h = \frac{1 - \frac{k}{k-1+e^\epsilon}}{1 + \frac{k}{k-1+e^\epsilon}}$, produces an estimate $\hat{q}$ that satisfies

$$p(y \in C_{\hat{q}}(x)) \geq 1 - \alpha - \Delta \quad \text{,w.p} \quad 1 - \delta$$

Theorem - LDP-CP-S

Local Differentially Private Conformal Prediction under Scores Privacy. Fix $\alpha, \delta, \Delta > 0$. There exists an ϵ-local differentially private algorithm that draws $n = O\left(\frac{T}{\Delta^2}\left(\frac{e^\epsilon+1}{e^\epsilon-1}\right)^2 \log(T/\delta)\right) \sim D$, produces an estimate $\hat{q}$ that satisfies

$$p(y \in C_{\hat{q}}(x)) \geq 1 - \alpha - \Delta \quad \text{,w.p} \quad 1 - \delta$$

Experiments

What we measure?

- size $= \frac{1}{n}\sum_i |C(x_i)|$, lower is better.
- coverage $= \frac{1}{n}\sum_i \mathbf{1}(y_i \in C(x_i))$, closer to $1-\alpha$ is better.
- Δ in $p(y \in C_{\hat{q}}(x)) \ \geq \ 1-\alpha-\Delta$, lower is better.

CP methods:

1. Not-Private-CP with coverage guarantee $1-\alpha$ - using a calibration set with clean labels
2. LDP-CP-{S,L} with coverage guarantee $1-\alpha-\Delta$
3. LDP-CP-{S,L}* with coverage guarantee $1-\alpha$

CP Results

Table: Calibration results for HPS and APS conformal scores across various datasets, using $\epsilon = 4$, $\epsilon^{\text{eff}} = \frac{\epsilon}{\sqrt{n}}$, and $\alpha = 0.1$ on 100 different seeds.

Dataset	**Method**	**HPS**		**APS**	
		size ↓	coverage (%)	size ↓	coverage (%)
OCTMNIST ($\epsilon^{\text{eff}} = 0.038$)	Not-Private-CP	2.57 ± 0.03	90.06 ± 0.99	2.61 ± 0.03	90.06 ± 0.97
	LDP-CP-L*	2.76 ± 0.04	92.22 ± 0.92	2.79 ± 0.03	92.28 ± 0.87
	LDP-CP-S*	2.97 ± 0.07	94.38 ± 0.81	2.99 ± 0.06	94.35 ± 0.70
TissueMNIST ($\epsilon^{\text{eff}} = 0.026$)	Not-Private-CP	5.55 ± 0.02	90.00 ± 0.24	5.58 ± 0.02	89.96 ± 0.24
	LDP-CP-L*	5.71 ± 0.02	91.68 ± 0.27	5.76 ± 0.02	91.70 ± 0.25
	LDP-CP-S*	6.12 ± 0.01	95.35 ± 0.07	5.83 ± 0.05	92.32 ± 0.45
OrganSMNIST ($\epsilon^{\text{eff}} = 0.080$)	Not-Private-CP	1.93 ± 0.05	90.09 ± 0.66	2.35 ± 0.05	90.10 ± 0.55
	LDP-CP-L*	2.77 ± 0.22	95.35 ± 0.74	3.35 ± 0.22	95.45 ± 0.74
	LDP-CP-S*	3.90 ± 0.03	97.75 ± 0.06	4.75 ± 0.03	98.40 ± 0.07
OrganAMNIST ($\epsilon^{\text{eff}} = 0.049$)	Not-Private-CP	1.19 ± 0.02	89.99 ± 0.46	1.61 ± 0.02	90.02 ± 0.38
	LDP-CP-L*	1.43 ± 0.03	93.62 ± 0.34	1.89 ± 0.05	93.51 ± 0.51
	LDP-CP-S*	1.88 ± 0.19	96.52 ± 0.82	2.44 ± 0.19	96.80 ± 0.60
OrganCMNIST ($\epsilon^{\text{eff}} = 0.081$)	Not-Private-CP	1.18 ± 0.03	89.99 ± 0.71	1.56 ± 0.03	90.02 ± 0.65
	LDP-CP-L*	1.63 ± 0.10	95.21 ± 0.74	2.05 ± 0.13	95.21 ± 0.80
	LDP-CP-S*	2.47 ± 0.02	98.18 ± 0.06	2.90 ± 0.04	98.15 ± 0.10

CP as a function of privacy ϵ

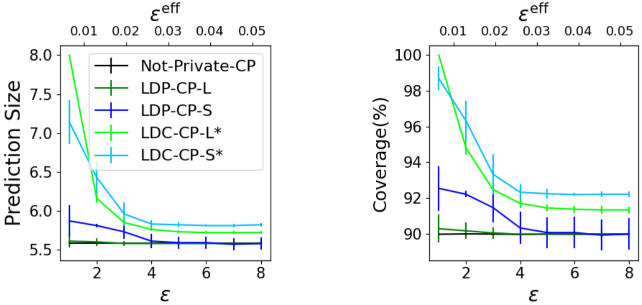


Figure: Size of prediction set (left) and coverage (right) as a function of the privacy ϵ (bottom x-axis) and effective privacy ϵ^{eff} (top x-axis). We show the (mean $\pm$ std) on TissueMNIST and APS score.

Conclusion and Open Questions

- We introduced two complementary Conformal Prediction (CP) approaches under Local Differential Privacy (LDP)
- LDP-CP-L: perturbs labels with randomized response + noise-aware calibration
- LDP-CP-S: users compute and perturb scores response locally
- Both methods ensure valid coverage guarantees while protecting user labels with ϵ-local DP

Thank You

Check out our full paper:
arxiv.org/abs/2505.15721
Code available:
github.com/cobypenso/local-differential-private-conformal-prediction

[DR13] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9, 01 2013.

[Dwo06] Cynthia Dwork. Differential privacy. In *International Colloquium on Automata, Languages, and Programming*, pages 1–12. Springer, 2006.

[PG24] Coby Penso and Jacob Goldberger. A conformal prediction score that is robust to label noise. In *MICCAI Int. Workshop on Machine Learning in Medical Imaging (MLMI)*, 2024.

[PGF25] Coby Penso, Jacob Goldberger, and Ethan Fetaya. Conformal prediction of classifiers with many classes based on noisy labels. In *Proceedings of the Symposium on Conformal and Probabilistic Prediction with Applications*, 2025.

[War65] Stanley L Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American statistical association*, 60(309):63–69, 1965.