EUREKA: A General Framework for Black-box Differential Privacy Estimators

Yun Lu
University of Victoria
yunlu@uvic.ca

Malik Magdon-Ismail
Rensselaer Polytechnic Institute
magdon@cs.rpi.edu

Yu Wei
Purdue University
yuwei@purdue.edu

Vassilis Zikas
Purdue University
vzikas@purdue.edu

Overview

Question
Differential privacy (DP) is a key tool in privacy-preserving data analysis. Yet it remains challenging for non-privacy-experts to prove the DP of their algorithms. Do we have a methodology for domain experts with limited data privacy background to empirically estimate the privacy of an arbitrary mechanism?

Our Results

• We showed the impossibility of the above task for unrestricted input domains, and introduce a natural, application-inspired relaxation of DP which we term relative DP.
• We proved a new link between the problems of DP parameter-estimation and Bayes optimal classifiers in ML.
• We propose a general framework for constructing and analyzing black-box DP estimators. The instantiated estimators achieve two desirable properties:
  • black-box, i.e., they do not require knowledge of the underlying mechanism
  • They have a theoretically-proven accuracy

Our Result in Details

What is relative DP?

Relative DP protects individual privacy \( x \) during the query \( M \) over dataset \( D \in T \).

Eureka Moment

A link casting the DP parameter-estimation problem to a binary classification problem.

For a binary classification problem with likelihood \( X, [Y]_1 \), uniform prior and 0/1 loss function, we have

\[
\delta_{X,Y}(\epsilon) = \max\left(1 - 2\exp(-\epsilon R(h^*)) - 2\epsilon\right),
\]

where \( R(h^*) \) is the risk of the optimal Bayes classifier for the problem, and \( \delta_{X,Y}(\epsilon) \) is defined as

\[
\delta_{X,Y}(\epsilon) = \max\left(\max_{h \in \mathcal{H}} \Pr[X \in S] - \epsilon \Pr[Y \in S], 0\right).
\]

\( [Y]_1 \) is a distribution for being a coin \( c \) where \( \Pr[c = 1] = \exp(\epsilon - \delta) \), outputting \( Y \) if \( c = 1 \) or \( 0 \) (a null value) otherwise.

Evaluation on a concrete estimator

Consider the set of mechanisms \( \mathcal{C} \subseteq \mathcal{H}^n \rightarrow \mathbb{R}^n \) whose output distribution has a density. Let \( T \subseteq \mathcal{X}^n \) be any set of databases with size less than \( n \). Let \( h \) be a kNN classifier which is constructed from \( n \) samples and \( k = \sqrt{n} \). Then there exists a \( n \) such that for all \( n > m \), the left-hand side algorithm is \( (\alpha, \beta) \)-Approximate Relative DP Estimator for \( C \), where \( \alpha = 2\epsilon \sqrt{\ln(8m)/\beta} \) and \( \beta = 2\epsilon \sqrt{\ln(8m)/\beta} \).

Theoretical-proven Accuracy

Eureka Framework: construct a estimator from any binary classifier

1. Given a tested mechanism \( M \), a binary classifier \( h \) and a database set \( T \), and a privacy parameter \( \epsilon \)
2. For any \( D_0, D_1 \in T \), do the following
   1. Set r.v. \( X = M(D_0), Y = M(D_1) \)
   2. Set binary classification problem with likelihood \( X, [Y]_1 \)
   uniform prior and 0/1 loss function
   3. Estimate the risk \( R(h) \) over this classification problem
   4. Use our link to convert the estimate in Step 3 to \( \delta_{D_0,D_1} \)
   3. Compute \( \delta^* = \max\{\delta_{D_0,D_1} \} \)
   4. Claim mechanism \( M \) satisfies \((\epsilon, \delta^*, T)\) relative DP.

Reference