Max-Information, Differential Privacy, and Post-Selection Hypothesis Testing

Ryan Rogers, Aaron Roth, Adam Smith, and Om Thakkar

Overall Goal: Maintain statistical validity in adaptive data analysis.

Adaptive Data Analysis
- Part of a line of work initiated by [DFH+15a, DFH+15b, HU14].
- In practice, data analysis is inherently interactive, where experiments may depend on previous outcomes from the same dataset.
- To allow the analyst to reuse the dataset for multiple experiments, we want to restrict the amount of information learned about the data so that later experiments are nearly independent of the data.

False Discovery
- We want to design valid hypothesis tests where probability of a false discovery < α.
- Design a test t and use a p-value to determine if model H₀ is likely given the data
  \[ p(a) = P_{X \sim H_0}(t(X) > a) \]
- Note that \( p(t(X)) \sim \text{Unif}(0,1) \). Rejecting H₀ if \( p(t(X)) < \alpha \) ensures false discovery is at most α.
- Framework crucially relies on test being chosen independently of the data.
- Has led to false discovery rates > α.

Valid p-Value Correction
- \( \gamma : [0,1] \rightarrow [0,1] \) is a valid p-value correction for selection procedure A if ∀α the following procedure has false discovery ≤ α:
  1. Select test \( t \leftarrow A(X) \)
  2. Reject H₀ if the p-value \( p(t(X)) \leq \gamma(\alpha) \)

Max-Information [DFH+15b]
- An algorithm A with bounded max-info allows the analyst to treat \( A(X) \) as if it is independent of data X up to a factor.
  \[ I^p_\infty(A(X),X) = \log \left( \sup_{\delta} \frac{P((A(X),X) \in O) - \beta}{P((A(X) \otimes X) \in O)} \right) \]
- Differentiate between product and general distributions
  \[ I^p_\infty(A; n) = \sup_{s \sim X \sim S} I^p_\infty(A(X),X) \]
  \[ I^p_{\infty,p}(A; n) = \sup_{s \sim X \sim S} I^p_\infty(A(X),X) \]
- \( I^p_{\infty,p}(A; n) \leq k \), leads to a p-value correction:
  \[ \gamma(\alpha) = (\alpha - \beta)/\beta \]

Algorithms with Bounded Max Info [DFH+15b]
- Algorithms A : \( D^n \rightarrow \mathcal{Y} \) with bounded max-info include:
  1. Pure \( \epsilon \)-differentially private algorithms
    \[ I^\epsilon_\infty(A; n) \leq \sqrt{n \log(1/\beta)} \]
  2. Bounded description length algorithms
    \[ I^\beta_{\infty,p}(A; n) \leq \log(|\mathcal{Y}|/\beta) \]

What About Max-Info for Approximate Differential Privacy?
- Best known algorithms for adaptive data analysis are approximate DP.
- Can we get better Max-Info bounds for \( (\epsilon, \delta) \)-DP.
- Huge improvement in using approximate differential privacy in composition: using \( T \) many \( \epsilon \)-DP algorithms leads to \( \epsilon T \)-DP but also \( (\epsilon \sqrt{T \log(1/\delta)}, \delta) \)-DP.

Positive Result
- If \( A : D^n \rightarrow \mathcal{Y} \) is \( (\epsilon, \delta) \)-DP then
  \[ I^\epsilon_{\infty,p}(A; n) \leq (\epsilon^2 + \sqrt{\epsilon \delta}) n \]

Product Distributions
- Nearly gives the tight generalization bounds of DP algorithms for low sensitive queries from [BNS+16], but does NOT apply to p-values.
- [RZ16] also give a method to correct p-values based on mutual info but we can get an improved correction factor via Max-Info.

Negative Result
- There exists and \( (\epsilon, \delta) \)-DP algorithm such that
  \[ I^\epsilon_{\infty,p}(A; n) \geq n - \log(1/\delta) \log(n)/\epsilon \]

General Distributions
- We know that Max-Info composes and so pure DP and bounded description length algorithms can be used in any order.
- Ordering matters: we prove the negative result by showing that composing a bounded description length algorithm followed by an approx-DP algorithms outputs the full dataset w.h.p.

Acknowledgements and References
- Supported in part by grants from the Sloan foundation and NSF grants: CNS-1253345, CNS-1513694, IIS-1447700.