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

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

Supported by grants from the Sloan Foundation and NSF: CNS-1253345, CNS-1513694, IIS-1447700.
Data Analysis

\[ X \sim p^n \]

Analysis \( A \)

\[ A(X) \]
Data Analysis

\[ X \sim p^n \]

Analysis \[ t \leftarrow f(A(X)) \]

Nothing Significant
A lot of existing theory assumes tests are selected *independently* of the data.
How can we provide statistically valid answers to adaptively chosen analyses?
How can we provide statistically valid answers to adaptively chosen analyses?
Adaptive Data Analysis

How can we provide statistically valid answers to adaptively chosen analyses?
Adaptive Data Analysis

Answer: Limit the info learned about the dataset [Dwork, Feldman, Hardt, Pitassi, Reingold, Roth’15].
Answer: Limit the info learned about the dataset [Dwork,Feldman,Hardt,Pitassi,Reingold,Roth’15].
Contributions

• Post-selection Hypothesis Testing
  • Bounded Max-Info $\Rightarrow$ Valid Tests
  • Tighter connection than previous results.

• Approximate Differential Privacy $\Rightarrow$ Bounded Max-Info
  • $k$ rounds of adaptivity: max-info $\sim k$ rather than $k^2$.

Generalizes and unifies previous work
Related Work

• Lots of work in statistics community on post-selection inference [Freedman’83], [Leeb,Potscher’06], [Berk,Brown,Buja,Zhang,Zhao’13], ...
  • Specific to type of analyses performed
• [DFHPRR](STOC’15,NIPS’15,Science’15)
  • Initial connections between information, privacy and adaptive analysis
• Accuracy for specific queries
  • [DFHPRR] (STOC’15,Science’15)
  • [Bassily,Nissim,Smith,Steinke,Stemmer,Ullman’16]
  • [Cummings,Ligett,Nissim,Roth,Wu’16]
  • [Russo,Zou’16]
  • [Wang,Lei,Fienberg’16]
• Impossibility results
  • [Hardt,Ullman’14], [Steinke,Ullman’15]
Hypothesis Testing

- Hypothesis test is defined by
  - null hypothesis $H_0 \subseteq \Delta(D)$ and
  - statistic:
    
    $$t : D^n \rightarrow \{\text{Inconclusive, Reject}\}$$

- A *False Discovery* is when $X \sim P^n$ and $P \in H_0$ but $t(X) = \text{Reject}$

- Classical results apply when $t$ is independent of $X$.
- Want to bound $\Pr[\text{False Discovery}]$ when $t \leftarrow A(X)$.
Max-Information [DFHPRR'15]

- Algorithm $A$ has small max-info
  $\Rightarrow A(X)$ and $X$ are “close” to independent.

- The $\beta$-approximate max-info between $A(X)$ and $X$ is

$$I_\infty^\beta(A(X); X) = \log \left( \sup_0 \frac{\Pr[(A(X), X) \in O] - \beta}{\Pr[(A(X'), X) \in O]} \right)$$
Max-Information of Algorithms \[DFHPRR'15\]

\[ I_\infty^\beta (A(X); X) = \log \left( \sup_\mathcal{O} \frac{\Pr[(A(X), X) \in \mathcal{O}]}{\Pr[(A(X'), X) \in \mathcal{O}]} - \beta \right) \]

An algorithm \( A \) has \( \beta \)-approximate max-info for data sets of size \( n \) if

\[
I_\infty^\beta (A; n) = \sup_{\mathcal{S}: X \sim \mathcal{S}} \left\{ I_\infty^\beta (A(X); X) \right\}
\]

\[
I_{\infty, \Pi}^\beta (A; n) = \sup_{P: X \sim P^n} \left\{ I_\infty^\beta (A(X); X) \right\}
\]

any data distribution restrict to product distribution
Post-selection Hypothesis Testing

• [RZ'16]: When mutual info $I(X; A(X))$ is bounded, we can control $\text{Pr[False Discovery]}$ for adaptively selected tests.
Post-selection Hypothesis Testing

- [RZ’16]: When mutual info $I(X; A(X))$ is bounded, we can control $\Pr[\text{False Discovery}]$ for adaptively selected tests.
- [This Paper]: We get a tighter connection via max-info

\[
\begin{align*}
\text{Bounded Mutual Info} & \quad \Rightarrow \quad \text{Bounded Max-Info} \\
I(X; A(X)) & \quad \Rightarrow \quad I_\infty^\beta (X; A(X)) \\
\Rightarrow & \quad \text{Bound on} \\
& \quad \Pr[\text{False Discovery}] \\
& \quad \text{when} \ t \leftarrow A(X)
\end{align*}
\]
What procedures $A$ have bounded max-info?

- [DFHPRR’15] Max-information bounds for:
  - Description Length – $\log(\text{image size of } A)$
Differential Privacy [Dwork, McSherry, Nissim, Smith’06]

A randomized algorithm $A: D^n \rightarrow Y$ is $(\varepsilon, \delta)$-differentially private if for any neighboring data sets $x, x' \in D^n$ and for any outcome $S \subseteq Y$ we have

$$P(A(x) \in S) \leq e^\varepsilon P(A(x') \in S) + \delta$$

If $\delta = 0$ we say pure DP, and otherwise approximate DP.
Technical Contributions

• [DFHPRR’15]: If $A: D^n \rightarrow T$ is $(\epsilon, 0)$-DP, then for $\beta > 0$,

$$I_{\infty, \Pi}^\beta (A; n) \leq \tilde{O}(\epsilon^2 n)$$

$$I_{\infty}^0 (A; n) \leq O(\epsilon n)$$

• [This paper]: If $A: D^n \rightarrow T$ is $(\epsilon, \delta)$-DP, then

$$I_{\infty, \Pi}^\beta (A; n) \leq \tilde{O}(\epsilon^2 n) \text{ where } \beta \approx n \sqrt{\frac{\delta}{\epsilon}}$$

• [This paper] (based on [De’12]): There exists an $(\epsilon, \delta)$-DP procedure $A$ where,

$$I_{\infty}^\beta (A; n) \approx n \text{ for any } \beta < \frac{1}{2} - \delta$$
Consequences of Positive Result

Theorem: If $A: D^n \rightarrow T$ is $(\epsilon, \delta)$-DP, then

$$I_{\infty, \Pi}^{\beta}(A; n) \leq \tilde{O}(\epsilon^2 n) \quad \text{where} \quad \beta \approx n \sqrt{\frac{\delta}{\epsilon}}$$

• Recover (optimal) results of [BNSSSU’16] for low sensitive queries.
  • However, our bounds apply more generally (e.g. adaptive hypothesis tests).

• Composition of $k$ adaptively selected $(\epsilon, 0)$-DP procedures: $A_1, \ldots, A_k$
  • [DFHPRR’15]: $I_{\infty, \Pi}^{\beta}(A_k \circ \cdots \circ A_1; n) \leq \tilde{O}(n \epsilon^2 k^2)$
  • [This Paper]: $I_{\infty, \Pi}^{\beta}(A_k \circ \cdots \circ A_1; n) \leq \tilde{O}(n \epsilon^2 k)$

Via strong composition theorem from [Dwork,Rothblum,Vadhan’10]
Contributions

• Post-selection Hypothesis Testing
  • Max-Info Bound $\implies$ bound $\Pr[\text{False Discovery}]$ in adaptive settings
  • Improves on previous result of [RZ’16] that uses mutual info.

• $(\epsilon, \delta)$-DP $\implies$ Bounded Max–Info over product distributions
  • Recovers results from [BNSSSU’16] that dealt with specific analyses.
  • $k$ rounds of adaptivity: we get max-info $\sim k$, where [DFHPRR’15] gives $\sim k^2$

Thanks!
Proof Sketch of Positive Result

Theorem: If $A: D^n \rightarrow T$ is $(\epsilon, \delta)$-DP, then

$$I_{\infty, \Pi}^\beta(A; n) \leq \tilde{O}(\epsilon^2 n) \text{ where } \beta \approx n \sqrt{\frac{\delta}{\epsilon}}$$

- Define the following random variable where $x \sim P^n, a \sim A(x)$ and

$$Z(a, x) = \log \left( \frac{\Pr[(A(X), X) = (a, x)]}{\Pr[(A(X'), X) = (a, x)]} \right)$$

$$= \sum_{i=1}^{n} \log \left( \frac{\Pr[X_i = x_i | a, x_1:i-1]}{\Pr[X_i = x_i]} \right) = \sum_{i=1}^{n} Z_i(a, x_1:i)$$

- Note that if we can bound this with high probability then we can bound approximate max-info.
Proof Sketch of Positive Result

• We want to apply a concentration bound (Azuma’s inequality) to the following quantity: $\sum_{i=1}^{n} Z_i(a, x_{1:i})$

• We must then have:
  • A bound on the expectation of each $Z_i(a, x_{1:i})$
  • A bound on each $Z_i(a, x_{1:i})$

• Problem: Each $Z_i(a, x_{1:i})$ is NOT bounded.

• Although each term is bounded with high probability, conditioning on the same $A(X) = a$ and a prefix of data $X_{1:i-1} = x_{1:i-1}$ in every term complicates the argument.
Proof Sketch of Positive Result

For any $t > 0$

$$\Pr\left[\sum_{i=1}^{n} Z_i(A, X_{1:i}) \geq \varepsilon^2 n + n\sqrt{\delta/\varepsilon} + t \varepsilon \sqrt{n}\right]$$

$$\leq \Pr\left[\sum_{i=1}^{n} Z_i(A, X_{1:i}) \geq \varepsilon^2 n + n\sqrt{\delta/\varepsilon} + t \varepsilon \sqrt{n} \cap (A, X) \in \text{GOOD}\right]$$

$$+ \Pr[(A, X) \in \text{BAD}]$$

$$\leq e^{\frac{-t^2}{2}} + O(n\sqrt{\delta/\varepsilon})$$

Set $t = O(\varepsilon \sqrt{n})$