Power of Ordered Hypothesis Testing

Lihua Lei and William Fithian

Department of Statistics, UC Berkeley

ICML 2016, New York
Table of Contents

1 Setup

2 Adaptive Seqstep

3 Power Comparison
Table of Contents

1 Setup

2 Adaptive Seqstep

3 Power Comparison
Multiple Testing Problem with FDR Control

- General setup: a sequence of hypotheses \( H_1, H_2, \ldots, H_n \);
- \( \mathcal{H}_0 = \{i : H_i \text{ is true}\} \) be the set of null hypotheses;
- \( S = \{i : H_i \text{ is rejected}\} \) be the set of discoveries;
- \( \text{FDP} = \frac{V}{R \lor 1} \) be the False Discovery Proportion with \( V = |S| \) and \( R = |S \cap \mathcal{H}_0| \);
- \( \text{FDR} = \mathbb{E}[\text{FDP}] \) be the False Discovery Rate, the target that a procedure should control.
- A procedure that control FDR at level 0.1 produces a rejection set \( S \) with roughly 90% being the true discoveries.
Ordered Hypothesis Testing

- Domain knowledge might be used to indicate which hypothesis is more “promising”, i.e. likely to be rejected;
- Heuristically, more focus should be put on “promising” hypotheses;
- Sort $H_1, \ldots, H_n$ from most “promising” to least “promising” via the prior knowledge;
- A procedure that takes advantage of the ordering is called an ordered hypothesis testing procedure.
GEOquery data\textsuperscript{1}[LB15] consists of gene expression measurements in response to estrogen in breast cancer cells; Consists of $n = 22283$ genes and two groups (a treatment group and a control group) with 5 trials in each; Test $H_i : F_{0i} = F_{1i}$, where $F_{0i}$ and $F_{1i}$ are the distributions of gene expression of gene $i$ in the control group and the treatment group, respectively; $H_1, \ldots, H_n$ are ordered by auxiliary data.

\footnotesize{\textsuperscript{1}http://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS2324}
Example: GEOquery Data

Ordered with High Dose Data

Ordered with Medium Dose Data

Lihua Lei and William Fithian
Power of Ordered Hypothesis Testing
Input: p-values $p_1, \ldots, p_n$.
Usually assume $p_i \sim U([0, 1])$ for $i \in \mathcal{H}_0$;
Input: p-values $p_1, \ldots, p_n$. Usually assume $p_i \sim U([0, 1])$ for $i \in \mathcal{H}_0$;

Rejection Set: a rectangular region indexed by $s$ and $k$;
Input: p-values $p_1, \ldots, p_n$. Usually assume $p_i \sim U([0, 1])$ for $i \in \mathcal{H}_0$;

Rejection Set: a rectangular region indexed by $s$ and $k$;

$\hat{\text{FDP}}$: an estimate of FDP for given rejection region;
Existing Methods Revisited: General Framework

- Input: p-values $p_1, \ldots, p_n$.
  Usually assume $p_i \sim U([0, 1])$ for $i \in \mathcal{H}_0$;

- Rejection Set: a rectangular region indexed by $s$ and $k$;

- $\hat{\text{FDP}}$: an estimate of FDP for given rejection region;

- Enlarge $k$ (fix $s$) as long as $\hat{\text{FDP}} \leq q$;

Lihua Lei and William Fithian
Power of Ordered Hypothesis Testing
Input: p-values $p_1, \ldots, p_n$. Usually assume $p_i \sim U([0, 1])$ for $i \in \mathcal{H}_0$;

Rejection Set: a rectangular region indexed by $s$ and $k$;

$\widehat{\text{FDP}}$: an estimate of FDP for given rejection region;

Enlarge $k$ (fix $s$) as long as $\widehat{\text{FDP}} \leq q$;

Reject all (red) points in the pink region.
Existing Methods Revisited: Accumulation Test

\[
\hat{FDP}_{\text{AT}} = \frac{C + \sum_{i=1}^{k} h(p_i)}{k + 1}
\]

- \( h \in [0, C], \int_0^1 h(x) \, dx = 1; \)
- Find the maximum \( k \) such that \( \hat{FDP}_{\text{AT}} \leq q; \)
- ForwardStop[GWCT15]:
  \[
  h(x) = -\log(1 - x);
  \]
- Seqstep[BC15]:
  \[
  h(x) = \frac{I(x > \lambda)}{1 - \lambda};
  \]
- HingeExp[LB15]:
  \[
  h(x) = -\frac{I(x > \lambda)}{1 - \lambda} \log\left(\frac{1 - x}{1 - \lambda}\right).
  \]
Existing Methods Revisited: Selective Seqstep

\[
\hat{\text{FDP}}_{SS} = \frac{ks}{R(k; s) \lor 1} \cdot \frac{A(k; s) + 1}{k(1 - s)}
\]

- \( R(k; s) = |\{i \leq k : p_i \leq s\}|; \)
- \( A(k; s) = |\{i \leq k : p_i > s\}|; \)
- \( s \) is pre-fixed;
- Find the maximum \( k \) such that \( \hat{\text{FDP}}_{SS} \leq q \).
- Turns out that the blue term should be an approximation of \( \pi_{0,k} \),
  \[
  \pi_{0,k} = \frac{|\{1, \ldots, k\} \cap \mathcal{H}_0|}{k};
  \]
- Too conservative for small \( s \).
Existing Methods Revisited: Adaptive Seqstep

$$\widehat{\text{FDP}}_{\text{AS}} = \frac{ks}{R(k; s) \lor 1} \cdot \frac{A(k; \lambda) + 1}{k(1 - \lambda)}$$

- $R(k; s) = |\{i \leq k : p_i \leq s\}|$;
- $A(k; \lambda) = |\{i \leq k : p_i > \lambda\}|$;
- $s$ and $\lambda$ are pre-fixed;
- Find the maximum $k$ such that $\widehat{\text{FDP}}_{\text{AS}} \leq q$;
- Much less conservative if a large $\lambda$, say 0.5, is used.
We prove that AS controls FDR in finite samples.

**Theorem 1.**

Assume that

1. \{p_i : i \in \mathcal{H}_0\} are independent of \{p_i : i \notin \mathcal{H}_0\};

2. \{p_i : i \in \mathcal{H}_0\} are i.i.d. with distribution function \(F_0\) that stochastically dominates \(U[0, 1]\).

Then AS controls FDR at level \(q\).
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Setup</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Adaptive Seqstep</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Power Comparison</td>
<td></td>
</tr>
</tbody>
</table>
Heuristic Comparison of Asymptotic Power: AS Versus AT

FDP of AT and AS (Good Ordering)

FDP of AT and AS (Bad Ordering)

AS is more robust to ordering than AT.
Utility Comparison of Asymptotic Power: AS Versus AT

FDP of AT and AS (Good Ordering)

- FDP of AT
- FDP of AS

FDP of AT and AS (Bad Ordering)

- FDP of AT
- FDP of AS

AS is more robust to ordering than AT.
FDPhat of SS and AS (Strong Signals)

FDPhat of SS and AS (Weak Signals)

AS is more robust to weak signals than SS.
Heuristic Comparison of Asymptotic Power: AS Versus SS

FDPhat of SS and AS (Strong Signals)

FDPhat of SS and AS (Weak Signals)

AS is more robust to weak signals than SS.
By analyzing the formulas of asymptotic power, we conclude that

<table>
<thead>
<tr>
<th></th>
<th>Weak Signals</th>
<th>Strong Signals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad Ordering</td>
<td>$\text{AS} \gg \text{SS}$, $\text{AS} \gg \text{AT}$</td>
<td>$\text{AS} &gt; \text{SS}$, $\text{AS} \gg \text{AT}$</td>
</tr>
<tr>
<td>Good Ordering</td>
<td>$\text{AS} \gg \text{SS}$, $\text{AS} &gt; \text{AT}$</td>
<td>$\text{AS} &gt; \text{SS}$, $\text{AT} \not&gt; \text{AS}$</td>
</tr>
</tbody>
</table>
Real Data Example: GEOquery Data

Ordered with High Dose Data

Target FDR level $q$

# of discoveries

0 500 1000 1500 2000 2500

AT (SeqStep, C=2)
AT (HingeExp, C=2)
AT (ForwardStop)
SS (s=q)
AS (s=q, lam=0.5)
BH
Storey

Ordered with Medium Dose Data

Target FDR level $q$

# of discoveries

0 500 1000 1500 2000 2500

AT (SeqStep, C=2)
AT (HingeExp, C=2)
AT (ForwardStop)
SS (s=q)
AS (s=q, lam=0.5)
BH
Storey

Lihua Lei and William Fithian

Power of Ordered Hypothesis Testing
Rina Foygel Barber and Emmanuel J Candès.
Controlling the false discovery rate via knockoffs.

Max Grazier G’Sell, Stefan Wager, Alexandra Chouldechova, and Robert Tibshirani.
Sequential selection procedures and false discovery rate control.

Ang Li and Rina Foygel Barber.
Accumulation tests for fdr control in ordered hypothesis testing.

John D Storey, Jonathan E Taylor, and David Siegmund.
Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach.
THANK YOU!