# Power of Ordered Hypothesis Testing

Lihua Lei and William Fithian

Department of Statistics, UC Berkeley

ICML 2016, New York

# Table of Contents

# Table of Contents

# 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;
- $\mathcal{S} = \{i : H_i \text{ is rejected}\}$ be the set of *discoveries*;
- $\mathrm{FDP} = \frac{V}{R \vee 1}$ be the *False Discovery Proportion* with $V = |\mathcal{S}|$ and $R = |\mathcal{S} \cap \mathcal{H}_0|$;
- $\mathrm{FDR} = \mathbb{E}\mathrm{FDP}$ 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 $\mathcal{S}$ with roughly 90% being the true discoveries.

# Ordered Hypothesis Testing

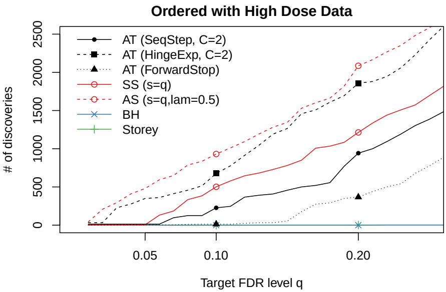
- Domain knowledge might be used to indicate which hypothesis is more "promising", i.e. likely to be rejected;
- Heuristically, more focus should be put on "promising" hypotheses;
- Sort $H_1, \ldots, H_n$ from most "promising" to least "promising" via the prior knowledge;
- A procedure that takes advantage of the ordering is called an *ordered hypothesis testing procedure*.

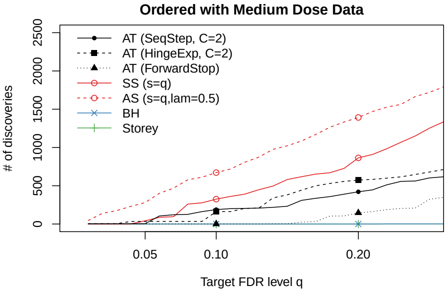
# Example: GEOquery Data

- GEOquery data[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.

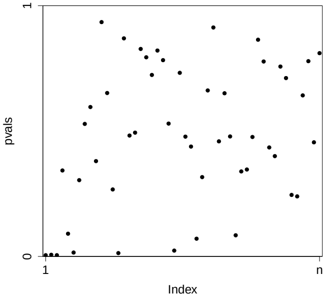
[1] http://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS2324

# Example: GEOquery Data



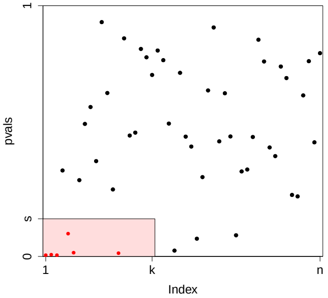



# Table of Contents



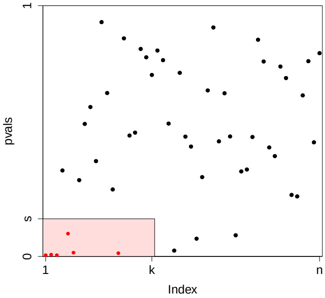

- Input: p-values $p_1, \ldots, p_n$. Usually assume $p_i \sim U([0,1])$ for $i \in \mathcal{H}_0$;

# Existing Methods Revisited: General Framework



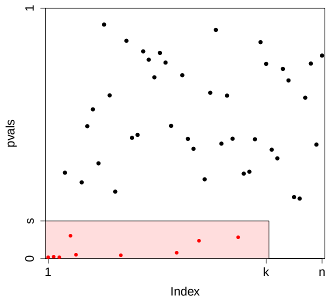

- Input: p-values $p_1, \ldots, p_n$. Usually assume $p_i \sim U([0,1])$ for $i \in \mathcal{H}_0$;
- Rejection Set: a rectangular region indexed by $s$ and $k$;

# Existing Methods Revisited: General Framework



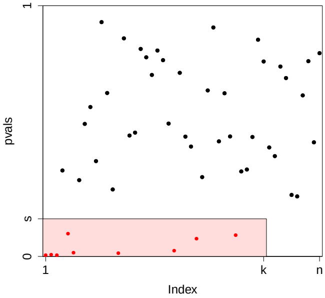

- Input: p-values $p_1, \ldots, p_n$. Usually assume $p_i \sim U([0,1])$ for $i \in \mathcal{H}_0$;
- Rejection Set: a rectangular region indexed by $s$ and $k$;
- $\widehat{\mathrm{FDP}}$: an estimate of $\mathrm{FDP}$ for given rejection region;

# Existing Methods Revisited: General Framework



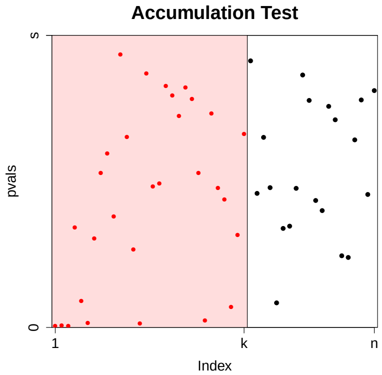

- Input: p-values $p_1, \ldots, p_n$. Usually assume $p_i \sim U([0,1])$ for $i \in \mathcal{H}_0$;
- Rejection Set: a rectangular region indexed by $s$ and $k$;
- $\widehat{\mathrm{FDP}}$: an estimate of $\mathrm{FDP}$ for given rejection region;
- Enlarge $k$ (fix $s$) as long as $\widehat{\mathrm{FDP}} \leq q$;

# 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 $k$;
- $\widehat{\mathrm{FDP}}$: an estimate of $\mathrm{FDP}$ for given rejection region;
- Enlarge $k$ (fix $s$) as long as $\widehat{\mathrm{FDP}} \leq q$;
- Reject all (red) points in the pink region.

$$\widehat{\mathrm{FDP}}_{AT} = \frac{C + \sum_{i=1}^{k} h(p_i)}{k+1}$$




- $h \in [0, C]$, $\int_0^1 h(x)dx = 1$;
- Find the maximum $k$ such that $\widehat{\mathrm{FDP}}_{AT} \le q$;
- ForwardStop[GWCT15]:

$$h(x) = -\log(1-x);$$

- Seqstep[BC15]:

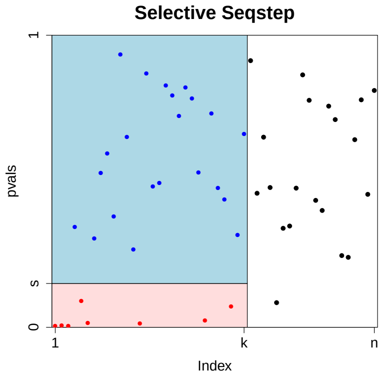
$$h(x) = \frac{I(x > \lambda)}{1-\lambda};$$

- HingeExp[LB15]:

$$h(x) = -\frac{I(x > \lambda)}{1-\lambda}\log\left(\frac{1-x}{1-\lambda}\right).$$

# Existing Methods Revisited: Selective Seqstep

$$\widehat{\mathrm{FDP}}_{SS} = \frac{ks}{R(k;s) \vee 1} \cdot \frac{A(k;s)+1}{k(1-s)}$$



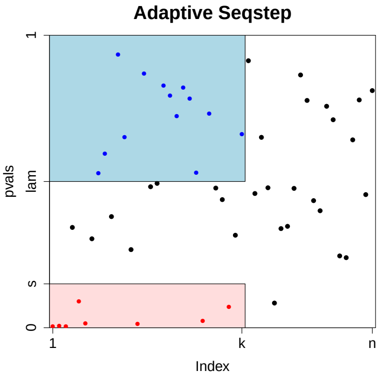

- $R(k;s) = |\{i \le k : p_i \le s\}|$;
- $A(k;s) = |\{i \le k : p_i > s\}|$;
- $s$ is pre-fixed;
- Find the maximum $k$ such that $\widehat{\mathrm{FDP}}_{SS} \le q$.
- Turns out that the blue term should be an approximation of $\pi_{0,k}$,

$$\pi_{0,k} = \frac{|\{1,\ldots,k\} \cap \mathcal{H}_0|}{k};$$

- Too conservative for small $s$.

# Existing Methods Revisited: Adaptive Seqstep

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




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

We prove that AS controls FDR in finite samples.

**Theorem 1.**

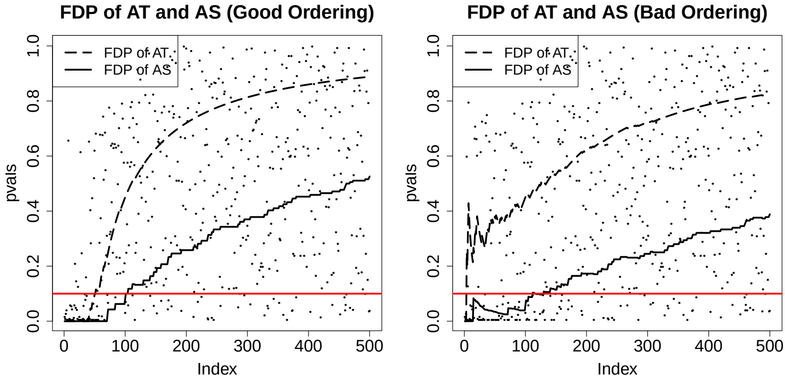
*Assume that*

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* $U[0, 1]$.

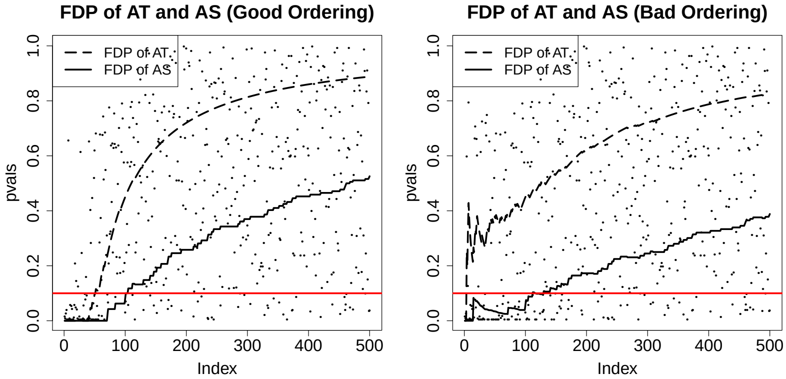
*Then AS controls FDR at level* $q$.

# Table of Contents

# Heuristic Comparison of Asymptotic Power: AS Versus AT

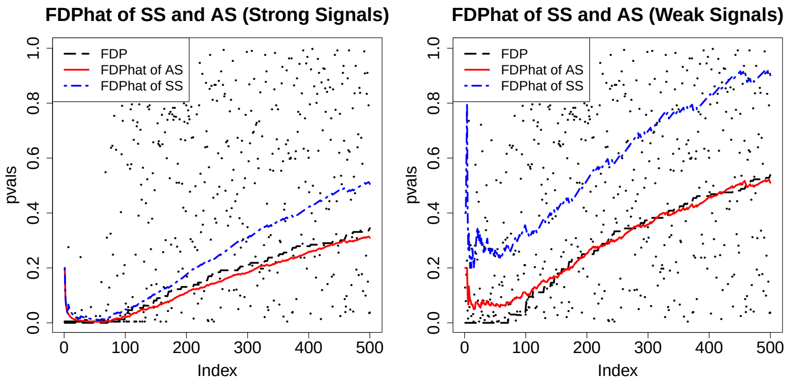


# Heuristic Comparison of Asymptotic Power: AS Versus AT

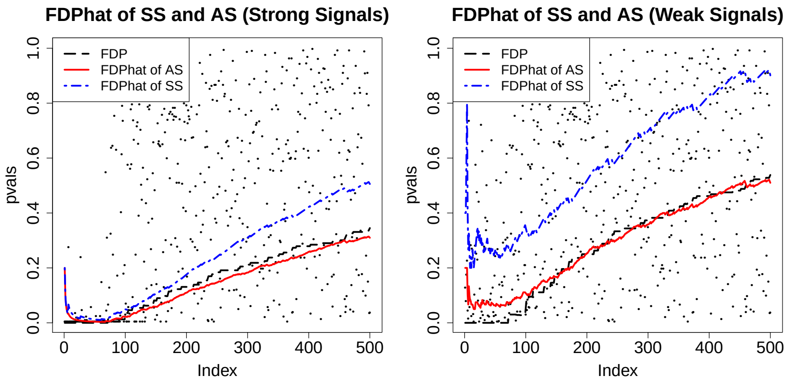



**AS is more robust to ordering than AT.**

# Heuristic Comparison of Asymptotic Power: AS Versus SS



# Heuristic Comparison of Asymptotic Power: AS Versus SS



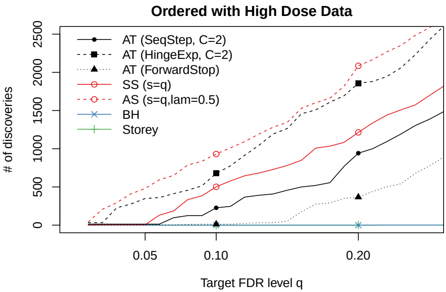

**AS is more robust to weak signals than SS.**

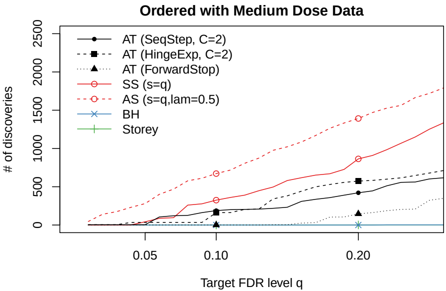
# Analytical Comparison of Asymptotic Power

By analyzing the formulas of asymptotic power, we conclude that

| | Weak Signals | Strong Signals |
|---|---|---|
| Bad Ordering | AS $\gg$ SS, AS $\gg$ AT | AS $>$ SS, AS $\gg$ AT |
| Good Ordering | AS $\gg$ SS, AS $>$ AT | AS $>$ SS, AT $\gtrless$ AS |

# Real Data Example: GEOquery Data



# References


Rina Foygel Barber and Emmanuel J Candès.
Controlling the false discovery rate via knockoffs.
*The Annals of Statistics*, 43(5):2055–2085, 2015.

Max Grazier G'Sell, Stefan Wager, Alexandra Chouldechova, and Robert Tibshirani.
Sequential selection procedures and false discovery rate control.
*Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2015.

Ang Li and Rina Foygel Barber.
Accumulation tests for fdr control in ordered hypothesis testing.
*arXiv preprint arXiv:1505.07352*, 2015.

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.
*Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(1):187–205, 2004.

# THANK YOU!