# Multiple Testing and Shrinkage Estimation

Debashis Ghosh

Department of Biostatistics

University of Michigan

Multiple Comparisons Procedures 2005

Joint work with Wei Chen and Trivellore Raghunathan

## Outline

I. Introduction

II. False discovery rate (FDR): definition

III. Model for FDR

IV. FDR and variable selection

V. Double shrinkage estimation

VI. Future Work

## Introduction

- Tremendous recent interest in multiple testing procedures
- Scientific application areas
  - Functional genomics
  - Brain imaging
  - Chemometrics
  - Astrophysics

## False discovery rate

- Suppose we are interested in testing a set of $m$ hypotheses
- For $m_0$ of them, the null is true

Table 1: Outcomes of $m$ tests of hypotheses

| | Accept | Reject | Total |
|---|---|---|---|
| True Null | U | V | $m_0$ |
| True Alternative | T | S | $m_1$ |
| | W | Q | $m$ |

- FDR defined as

$$FDR \equiv E\left[\frac{V}{Q} \mid Q > 0\right] P(Q > 0).$$

## Multiple testing: philosophy

- FDR is more liberal than familywise error rate for certain situations
- Goal with massively multiple testing problems: selection
- Want to make selections that have a high probability of being "real discoveries"
- Thus, what is really important are *correctly calibrated* inferences
- Bayesian (and more generally shrinkage) approaches offer such a calibration

## Goals of research

- Study FDR behavior from a risk point of view
- Relate FDR to variable selection procedures
- Propose shrinkage estimators for multiple testing

## FDR: mixture model

- Let $T_1, \ldots, T_m$ be independent test statistics
- Let $H_1, \ldots, H_m$ be indicator variables where $H_i = 0$ if the $i$th null hypothesis is true and $H_i = 1$ if the $i$th alternative hypothesis is true.
- $H_1, \ldots, H_m$ are a random sample from a Bernoulli distribution where for $i = 1, \ldots, m$, $P(H_i = 0) = \pi_0$
- Storey (2002) proved that

$$\begin{aligned} pFDR(R) &= P(H = 0 | T \in R) \\ &= \frac{\pi_0 P(T \in R | H = 0)}{P(T \in R)}, \end{aligned}$$

where $pFDR = E\left[\frac{V}{Q} \mid Q > 0\right]$

## FDR: mixture model (cont'd.)

- Note that pFDR does not condition on all data
- Local FDR, defined as $P(H = 0|T_1, \ldots, T_m)$ is fully conditional
- Bias-variance tradeoff in choice of cardinality of data points to condition on
- Test statistics get "used" as data points

## FDR and variable selection

- Assume probabilistic framework of George and McCulloch for predictor selection in regression

$$
\begin{aligned}
Y_i &\overset{ind}{\sim} N(\mathbf{X}_i^T\beta, \sigma^2) && (1)\\
\beta_i|\gamma_i &\sim (1-\gamma_i)N(0,\tau_i^2) + \gamma_i N(0, c_i^2\tau_i^2) && (2)\\
\gamma_i &\overset{ind}{\sim} Be(p_i) && (3)\\
\sigma^2 &\sim IG(\nu/2, \nu/2) && (4)
\end{aligned}
$$

- Rank based on the posterior distribution of $\gamma_i$
- Then the local FDR at zero is

$$P(\gamma_i = 0|\hat{\beta}_i = 0)$$

- The false discovery rate based on $\hat{\beta}_l$ being in a critical region $R$ is

$$FDR(R) \equiv \frac{\int_{x\in R}\{2\pi(\sigma_l^2 + c_l^2\tau_l^2)\}^{-1/2}\exp\{-x^2/(\sigma_l^2 + c_l^2\tau_l^2)\}dx}{\int_{x\in R}\{2\pi(\sigma_l^2 + \tau_l^2)\}^{-1/2}\exp\{-x^2/(\sigma_l^2 + \tau_l^2)\}dx}.$$

## Proposed variable selection procedure

(a) Set level to be $\alpha$ and fix a rejection region $R$.

(b) Fit model (1) - (4) using Markov Chain Monte Carlo (MCMC) methods.

(c) Based on the MCMC output, calculate $pp_i \equiv P(\gamma_i = 0 | \hat{\beta}_i \in R)$, $i = 1, \ldots, G$.

(d) Let $pp_{(1)} \leq pp_{(2)} \leq \cdots \leq pp_{(G)}$ denote the sorted values of $pp_1, \ldots, pp_n$ in increasing order.

(e) Find $\hat{k} = \max\{1 \leq k \leq G : pp_{(k)} \leq \alpha k/G\}$; select variables $1, \ldots, G$.

## FDR mixture model: revisited

- We can think of the mixture model for testing as defining two estimation targets
- Consider shrinkage estimation in this setting
- **Note:** Shrinkage will only occur if test statistics have differing variances under null and alternative

## Double shrinkage estimators

- Shrink test statistics towards two targets corresponding to null and alternative hypotheses, $\mu_0$ and $\mu_1$
- Assume $\pi_0$ is known
- With respect to the first component, a shrinkage estimator is given by

$$T_{0i}^{JS} = T_i - \left[1 \wedge \frac{n-2}{\sum_{i=1}^{n}(T_i - \mu_0)^2}\right](T_i - \mu_0), \tag{5}$$

- For the second component,

$$T_{1i}^{JS} = T_i - \left[1 \wedge \frac{n-2}{\sum_{i=1}^{n}(T_i - \mu_1)^2}\right](T_i - \mu_1) \tag{6}$$

- A shrinkage estimator combining (5) and (6) is then given by $T_i^{JS} = \pi_0(T_i)T_{0i}^{JS} + \pi_1(T_i)T_{1i}^{JS}$, $i = 1, \ldots, n$, where

$$\pi_k(T_i) = \frac{\pi_k f_k(T_i)}{\pi_0 f_0(T_i) + \pi_1 f_1(T_i)}. \tag{7}$$

and $f_0$ and $f_1$ refer to the marginal densities of the distribution of the test statistics under the null and alternative hypotheses

## Double shrinkage estimators (cont'd.)

- This can be done with p-values as well (SPADE)
- Issues:
  1. Estimating $\pi_0$ from data
  2. What does a p-value estimate?
- This gives correctly calibrated measures of evidence that adjusts for multiple testing

## Simulation results

Table 2: Estimated mean-squared errors from simulation studies

| Effect | $\pi_0$ | Q-value | SPADE1 | SPADE2 | SPADE3 |
|---|---|---|---|---|---|
| Small | 0.2 | 0.179 | 0.186 | 0.180 | 0.16 |
| | 0.5 | 0.264 | 0.333 | 0.358 | 0.302 |
| | 0.8 | 0.165 | 0.333 | 0.380 | 0.31 |
| Medium | 0.2 | 0.179 | 0.183 | 0.171 | 0.168 |
| | 0.5 | 0.272 | 0.328 | 0.326 | 0.309 |
| | 0.8 | 0.168 | 0.330 | 0.374 | 0.319 |
| Large | 0.2 | 0.161 | 0.166 | 0.173 | 0.164 |
| | 0.5 | 0.251 | 0.297 | 0.275 | 0.296 |
| | 0.8 | 0.161 | 0.312 | 0.310 | 0.312 |

**Note**: Q-value refers to method of Storey and Tibshirani (2003). SPADE1 is the SPADE methodology, where $\pi_0$ is estimated using algorithm of Storey and Tibshirani (2003); SPADE2 is based on Pounds and Cheng (2004) method for estimation of $\pi_0$; SPADE3 is based on Dalmasso et al. (2005) method for estimation of $\pi_0$.

## Gene expression example

- Differential expression analysis focusing on localized versus metastatic prostate cancer
- 59 localized samples and 20 metastatic samples
- Following preprocessing steps:
  1. Genes that were reported as missing in more than 10% of samples were filtered out.
  2. Genes that had a sample variation greater than 0.15 across all samples
- Total of $m = 5241$ genes
- p-values based on $N(0, 1)$ distribution for t-statistic

## Gene expression example (cont'd.)

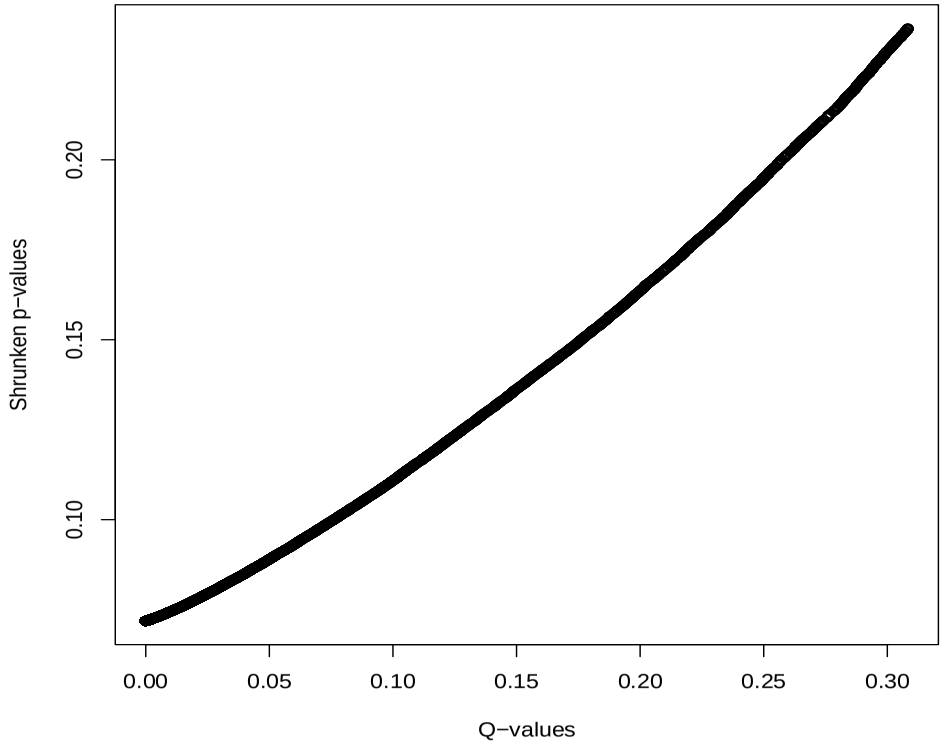


Figure 1: Plot of q-values using Storey (2002) method (horizontal axis) versus shrunken p-values from SPADE.

## Discussion

- In many multiple testing problems, what matters is having *calibrated* inferences
- Shrinkage/Bayesian approaches achieve this objective
- Future work:
  1. Synthesized framework using Bayes factors
  2. Graduated differential expression

## References


- Ghosh, D., Chen, W. and Raghunathan, T. E. (2004). The false discovery rate: a variable selection perspective.
  http://www.bepress.com/umichbiostat/paper41/
- Ghosh, D. (2005). Simultaneous estimation procedures and multiple testing: a decision-theoretic framework.
  http://www.bepress.com/umichbiostat/paper54/
- Ghosh, D. (2005). Shrunken p-values for assessing differential expression, with applications to genomic data analysis.
  http://www.bepress.com/umichbiostat/paper55/