Statistical Multiple Decision Making

Edsel A. Peña
University of South Carolina
Columbia, South Carolina
(E-Mail: pena@stat.sc.edu)

University of Miami Colloquium Talk
April 22, 2011

Outline

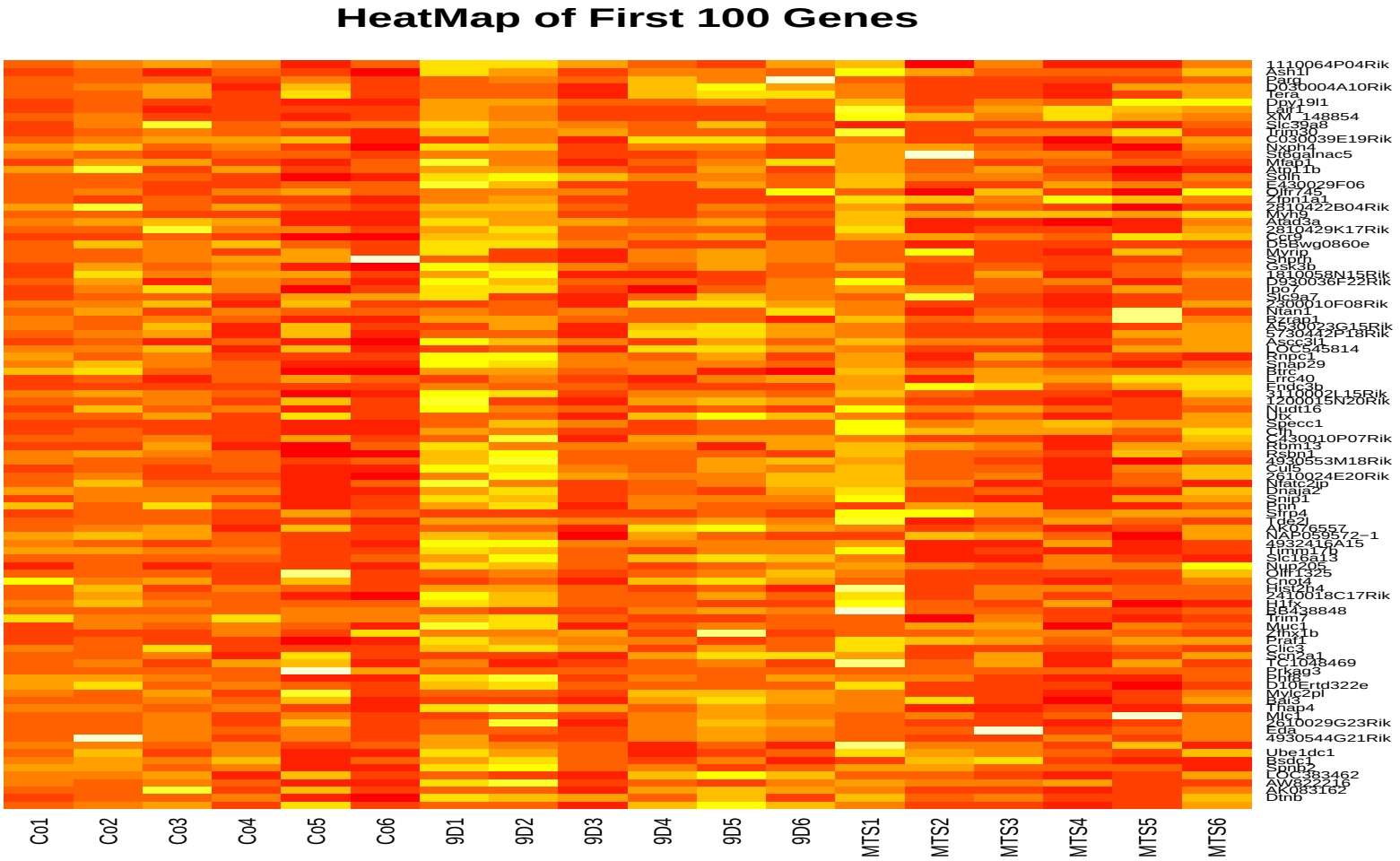
- Some Motivating Problems.
- Multiple Decision Problems.
- Mathematical Framework (Decision Functions, Losses, Risks).
- Special Case: Optimal Choice Between Two Actions.
- Multiple Decision Processes.
- Multiple Decision Size Function.
- Class of FWER-Controlling MDFs.
- Class of FDR-Controlling MDFs.
- An Application to a Microarray Data Set.
- Towards Optimal MDFs.
- Applicability and Some Comparisons.

Some Motivating Questions and Areas of Relevance

- Microarray data analysis: Which genes are relevant?
- Variable selection: Which of many predictors are relevant?
- Survival analysis: Which predictors affects a lifetime variable?
- Reliability: Which components in a system are relevant?
- Epidemiology: Spread of a disease in a geographical area.
- Oil (mineral) exploration: Where to dig?
- Business: Locations of business ventures.
- Sporting Events: Predicting outcomes of NBA playoff games.

A Microarray Data: HeatMap of Gene Expression Levels

First 100 genes out of 41267 genes in a colon cancer study at USC (M Peña's Lab). Three groups (Control; 9 Days; 2 Weeks) with 6 replicates each.



A Typical Variable Selection Problem

- Model.

$$Y = \beta_0 + \sum_{j=1}^{M} \beta_j X_j + \epsilon$$

- M is large, but many β_js are equal to zero.
- Observed Data: For $j = 1, 2, \ldots, n$,

$$(Z_j, \delta_j, X_{1j}, X_{2j}, \ldots, X_{Mj})$$

with

$$Z_j = \min(Y_j, C_j) \quad \text{and} \quad \delta_j = I\{Y_j \leq C_j\}$$

- Goal: To select the relevant predictor variables.

A Reliability (or Biological Pathways) Problem

- System is composed of components.
- Structure function, ϕ, relates components to system: series, parallel, series-parallel, etc.
- M potential components that could constitute a system. We do not know which components are relevant nor do we know the structure function.
- Question: Given data regarding the states or lifetimes of the system and components, how could we determine which components are relevant for this system?
- Component lifetimes may be censored by system lifetime.
- Highly nonlinear types of relationships.

The General Decision Problem

- We would like to discover the value of a parameter

$$\theta = (\theta_1, \theta_2, \ldots, \theta_M) \in \Theta = \{0, 1\}^M$$

- $\theta_m = 1$ means mth component is relevant; $\theta_m = 0$ means mth component is not relevant.
- Want to choose an action

$$a = (a_1, a_2, \ldots, a_M) \in \mathfrak{A} = \{0, 1\}^M$$

- $a_m = 1$ means we declare that $\theta_m = 1$, called a discovery; $a_m = 0$ means we declare that $\theta_m = 0$, a non-discovery.

Assessing our Actions: Losses

- Family-wise error indicator:

$$L_0(a,\theta) = I\left\{\sum_{m=1}^{M} a_m(1-\theta_m) > 0\right\}$$

- False Discovery Proportion:

$$L_1(a,\theta) = \frac{\sum_{m=1}^{M} a_m(1-\theta_m)}{\max\{\sum_{m=1}^{M} a_m, 1\}}$$

- Missed Discovery Proportion:

$$L_2(a,\theta) = \frac{\sum_{m=1}^{M} (1-a_m)\theta_m}{\max\{\sum_{m=1}^{M} \theta_m, 1\}}$$

If Only We Still Have **Paul, the Oracle**!

Sadly (or, Gladly), Revert to Being Statisticians!

- Obtain a BIG data (e.g., microarrays, Netflix):

$$X \in \mathfrak{X}$$

- Probabilistic Structure:

$$X \sim P$$

- Marginal Components:

$$X_m = z_m(X) \in \mathfrak{X}_m \quad \text{and} \quad X_m \sim P_m = P z_m^{-1}$$

- Parameters of Interest:

$$\theta_m = \theta_m(P_m)$$

- Example:

$$\theta_m = 1 \Longleftrightarrow P_m \in \{N(\mu, \sigma^2) : \mu \geq 0, \sigma^2 > 0\}$$

Multiple Decision Functions

- Multiple Decision Function:

$$\delta : \mathfrak{X} \to \mathfrak{A}$$

- Components:

$$\delta = (\delta_1, \delta_2, \ldots, \delta_M)$$

$$\delta_m : \mathfrak{X} \to \{0, 1\}$$

- $\mathfrak{D}$: space of multiple decision functions.
- $\mathcal{M}_0 = \{m : \theta_m = 0\}$ and $\mathcal{M}_1 = \{m : \theta_m = 1\}$
- Structure: $\{\delta_m(X) : m \in \mathcal{M}_0\}$ is an independent collection, and is independent of $\{\delta_m(X) : m \in \mathcal{M}_1\}$.
- $\{\delta_m(X) : m \in \mathcal{M}_1\}$ need NOT be an independent collection.

Risk Functions: Averaged Losses

- Given a $\delta \in \mathfrak{D}$:
- Family-Wise Error Rate (FWER):

$$R_0(\delta, P) = E[L_0(\delta(X), \theta(P))]$$

- False Discovery Rate (FDR):

$$R_1(\delta, P) = E[L_1(\delta(X), \theta(P))]$$

- Missed Discovery Rate (MDR):

$$R_2(\delta, P) = E[L_2(\delta(X), \theta(P))]$$

- Expectations are with respect to $X \sim P$.
- Goal: Choose $\delta \in \mathfrak{D}$ with small risks, whatever P is.

Special Case: A Pair of Choices ($M = 1$)

- $\theta \in \Theta = \{0, 1\}$
- $a \in \mathfrak{A} = \{0, 1\}$
- $L_0(a, \theta) = L_1(a, \theta) = aI(\theta = 0)$
- $L_2(a, \theta) = (1 - a)I(\theta = 1)$
- $X \sim P$ with $P \in \{P_0, P_1\}$

- $R_0(\delta, \theta) = R_1(\delta, \theta) = P_0(\delta(X) = 1)I(\theta = 0)$
- $R_2(\delta, \theta) = [1 - P_1(\delta(X) = 1)]I(\theta = 1)$

- Assume P_0 and P_1 have respective densities:

$$f_0(x) \quad \text{and} \quad f_1(x)$$

Types I and II Errors, Power, and Optimality

- $R_0(\delta, \theta)$: Type I error probability.
- $R_2(\delta, \theta)$: Type II error probability.
- Note

$$R_2(\delta, \theta = 1) = 1 - \pi(\delta)$$

where

$$\pi(\delta) = P_1(\delta(X) = 1) = \text{POWER of } \delta.$$

- Desired Goal: Given Type I level $\alpha \in [0, 1]$, find $\delta^*(\cdot; \alpha)$ with

$$R_0(\delta^*, \theta) \leq \alpha, \quad \text{for all } \theta,$$

and

$$R_1(\delta^*, \theta) \leq R_1(\delta, \theta), \quad \text{for all } \theta,$$

for any other δ with $R_1(\delta, \theta) \leq \alpha, \forall \theta$.

Neyman-Pearson MP Test δ^*_α

- Neyman and Pearson (1933) obtained the optimal [most powerful] decision function to be of form

$$\delta^*_\alpha(x) = \begin{cases} 1 & \text{if} \quad f_1(x) > c(\alpha) f_0(x) \\ \gamma(x) & \text{if} \quad f_1(x) = c(\alpha) f_0(x) \\ 0 & \text{if} \quad f_1(x) < c(\alpha) f_0(x) \end{cases}$$

  where $c(\alpha)$ and $\gamma(\alpha)$ satisfy

$$R_0(\delta^*_\alpha, \theta = 0) = \alpha.$$

- Remark: Depends on α, hence power depends on α.
- Leads to the notion of a decision process.

Concrete Example of a Decision Process

- Model: $X = (X_1, X_2, \ldots, X_n) \overset{IID}{\sim} N(\mu, \sigma^2)$.
- Problem: Test $H_0 : \mu \leq \mu_0$ $[\theta = 0]$ vs $H_1 : \mu > \mu_0$ $[\theta = 1]$
- Decision Function: t-test of size α given by

$$\delta(X; \alpha) = I\left\{\frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \geq t_{n-1;\alpha}\right\}$$

- Decision function depends on the size index α.
- Decision Process:

$$\Delta = (\delta(\alpha) \equiv \delta(\cdot; \alpha) : \alpha \in [0, 1])$$

- Size Condition:

$$\sup\{E_P[\delta(X; \alpha)] : \theta(P) = 0\} \leq \alpha$$

Multiple Decision Process

- Consider a multiple decision problem with M components.
- Multiple Decision Process:

$$\boldsymbol{\Delta} = (\Delta_m : m \in \mathcal{M} = \{1, 2, \ldots, M\})$$

- Decision Process for mth Component:

$$\Delta_m = (\delta_m(\alpha) : \alpha \in [0, 1])$$

- Example: t-test decision process for each component.
- Usual Approach: Pick a δ_m from Δ_m using the same α.
- Common Choices for α: (weak) FWER Threshold of q use:

$$\text{Bonferroni:} \quad \alpha = q/M$$

$$\text{Sidak:} \quad \alpha = 1 - (1 - q)^{1/M}$$

Notion of Size Functions

- A size function is a function

$$A : [0, 1] \to [0, 1]$$

  which is continuous, strictly increasing, $A(0) = 0$ and $A(1) \leq 1$, and possibly differentiable.

- Bonferroni size function: $A(\alpha) = \alpha/M$
- Sidak size function: $A(\alpha) = 1 - (1 - \alpha)^{1/M}$

- $\mathfrak{S}$: collection of possible size functions.
- Given a decision process Δ and a size function A, we choose the decision function from Δ according to

$$\delta[A(\alpha)].$$

Multiple Decision Size Function

- For a multiple decision problem with M components, a multiple decision size function is

$$\mathbf{A} = (A_m : m \in \mathcal{M}) \quad \text{with} \quad A_m \in \mathfrak{S}.$$

- **Condition:**

$$1 - \prod_{m \in \mathcal{M}} [1 - A_m(\alpha)] \leq \alpha$$

- Given a $\Delta = (\Delta_m : m \in \mathcal{M})$ and an $\mathbf{A} = (A_m : m \in \mathcal{M})$, multiple decision function is chosen according to

$$\delta(\alpha) = (\delta_m[A_m(\alpha)] : m \in \mathcal{M})$$

- Weak FWER of $\delta(\alpha)$:

$$R_0(\delta(\alpha), P) \quad = \quad 1 - \prod [1 - A_m(\alpha)] \leq \alpha$$

Neyman-Pearson Paradigm

- Control Type I error rate; minimize Type II error rate.
- Desired Type I error threshold: $q \in (0, 1)$
- Weak Control: For P with $\theta_m(P) = 0$ for all m, want a δ with

$$R_0(\delta, P) \leq q \quad \text{or} \quad R_1(\delta, P) \leq q.$$

- Strong Control: Whatever P is, want a δ such that

$$R_0(\delta, P) \leq q \quad \text{or} \quad R_1(\delta, P) \leq q.$$

- And, if above Type I error control is achieved, we want to have $R_2(\delta, P)$ small, if not optimal.

Towards Strong FWER Control

Given a MDP $\Delta = (\Delta_m)$ and MDS $\mathbf{A} = (A_m)$, for the chosen δ at α, its FWER is

$$\begin{aligned} R_0(\delta, P) &= E_P\left\{ I\left(\sum \delta_m[A_m(\alpha)][1-\theta_m(P)] > 0 \right) \right\} \\ &= P\left\{ \sum_{\mathcal{M}_0} \delta_m[A_m(\alpha)] > 0 \right\} \\ &= 1 - \prod_{\mathcal{M}_0} [1 - A_m(\alpha)] \\ &= 1 - \prod [1 - A_m(\alpha)]^{1-\theta_m(P)} \end{aligned}$$

Question: Given a threshold of q, what is the best α?

'Best' Choice of α

- **Oracle Paul**'s Choice:

$$\alpha^{\dagger}(q; P) = \inf\left\{\alpha \in [0,1] : \prod [1 - A_m(\alpha)]^{1-\theta_m(P)} < 1 - q\right\}$$

- But, P is unknown, hence $\theta_m(P)$ is also unknown. But we could estimate $\theta_m(P)$ by

$$\delta_m[A_m(\alpha)-].$$

- The Oracle's choice is then estimated by

$$\alpha^{\dagger}(q) = \inf\left\{\alpha \in [0,1] : \prod [1 - A_m(\alpha)]^{1-\delta_m[A_m(\alpha)-]} < 1 - q\right\}$$

Strong FWER-Controlling MDF

- Chosen Multiple Decision Function:

$$\delta^{\dagger}(q) = \left(\delta_m[A_m(\alpha^{\dagger}(q))] : m \in \mathcal{M}\right)$$

- Theorem

Given a $\Delta = (\Delta_m)$ *and an* $\mathbf{A} = (A_m)$*, the* $\delta^{\dagger}(q)$ *defined above has*

$$R_0(\delta^{\dagger}(q), P) \leq q$$

whatever P *is. That is, it is an MDF achieving strong FWER control at level* q*.*

Generalized P-Values

- Definition

  The mth component of the vector of generalized P-value statistic associated with Δ and $\mathbf{A}$ is

$$\alpha_m \equiv \alpha_m(\Delta, \mathbf{A}) = \inf\{\alpha \in [0,1] : \ \delta_m[A_m(\alpha)] = 1\}$$

  - Smallest size to decide in favor of $\theta_m = 1$ under $(\Delta, \mathbf{A})$.
  - Ordered Generalized P-Value Statistics:

$$0 \equiv \alpha_{(0)} < \alpha_{(1)} < \alpha(2) < \ldots < \alpha_{(M)} < \alpha_{(M+1)} \equiv 1$$

  - Observe that for

$$\alpha \in [\alpha_{(k)}, \alpha_{(k+1)}) \Longleftrightarrow \sum \delta_m[A_m(\alpha)] = k$$

Towards FDR Control

- Given MDP $\Delta = (\Delta_m)$ and MDS $\mathbf{A} = (A_m)$, the MDF

$$\delta(\alpha) = (\delta_m[A_m(\alpha)] : m \in \mathcal{M})$$

has FDR

$$R_1(\delta(\alpha), P) = E_P \left\{ \frac{\sum \delta_m[A_m(\alpha)](1 - \theta_m(P))}{\sum \delta_m[A_m(\alpha)]} \right\}$$

- Observe:

$$E_P \left\{ \sum \delta_m[A_m(\alpha)](1 - \theta_m(P)) \right\} \leq \sum A_m(\alpha)$$

'Best' Choice of α

- Preceding considerations heuristically suggest the α:

$$\alpha^*(q) = \sup\left\{\alpha \in [0,1] : \sum A_m(\alpha) \leq q \sum \delta_m[A_m(\alpha)]\right\}$$

- Chosen Multiple Decision Function:

$$\delta^*(q) = (\delta_m[A_m(\alpha^*(q))] : m \in \mathcal{M})$$

- Theorem

  Given a pair $(\Delta, \mathbf{A})$, *the MDF* $\delta^*(q)$ *achieves FDR control at level* q *in that*

$$R_1(\delta^*(q), P) \leq q.$$

Classes of MDFs Controlling FWER and FDR

- A class of strong FWER-controlling MDFs at threshold q is:

$$\mathfrak{D}^{\dagger} = \left\{ \delta^{\dagger}(q; \Delta, \mathbf{A}) : \Delta \in \mathfrak{D}, \mathbf{A} \in \mathfrak{S} \right\}$$

- A class of FDR-controlling MDFs at threshold q is:

$$\mathfrak{D}^{*} = \{ \delta^{*}(q; \Delta, \mathbf{A}) : \Delta \in \mathfrak{D}, \mathbf{A} \in \mathfrak{S} \}$$

- Remark: Sidak's sequential step-down strong FWER controlling MDF belongs to $\mathfrak{D}^{\dagger}$.
- Remark: Benjamini-Hochberg's step-up FDR controlling MDF belongs to $\mathfrak{D}^{*}$.
- Potential Utility: May choose best MDF in $\mathfrak{D}^{\dagger}$ or $\mathfrak{D}^{*}$ wrt the missed discovery rate.

Recalling BH FDR-Controlling MDF

- Benjamini-Hochberg (JRSS B, '95) paper. Most well-known FDR-controlling procedure.
- Let $P_1, P_2, \ldots, P_M$ be the ordinary P-values from the M tests.
- Let $P_{(1)} < P_{(2)} < \ldots < P_{(M)}$ be the ordered P-values.
- For FDR-threshold equal to q, define

$$K = \max \left\{ k \in \{0, 1, 2, \ldots, M\} : \; P_{(k)} \leq \frac{qk}{M} \right\}.$$

- BH MDF $\delta^{BH}(q) = (\delta_m^{BH} : m \in \mathcal{M})$ has

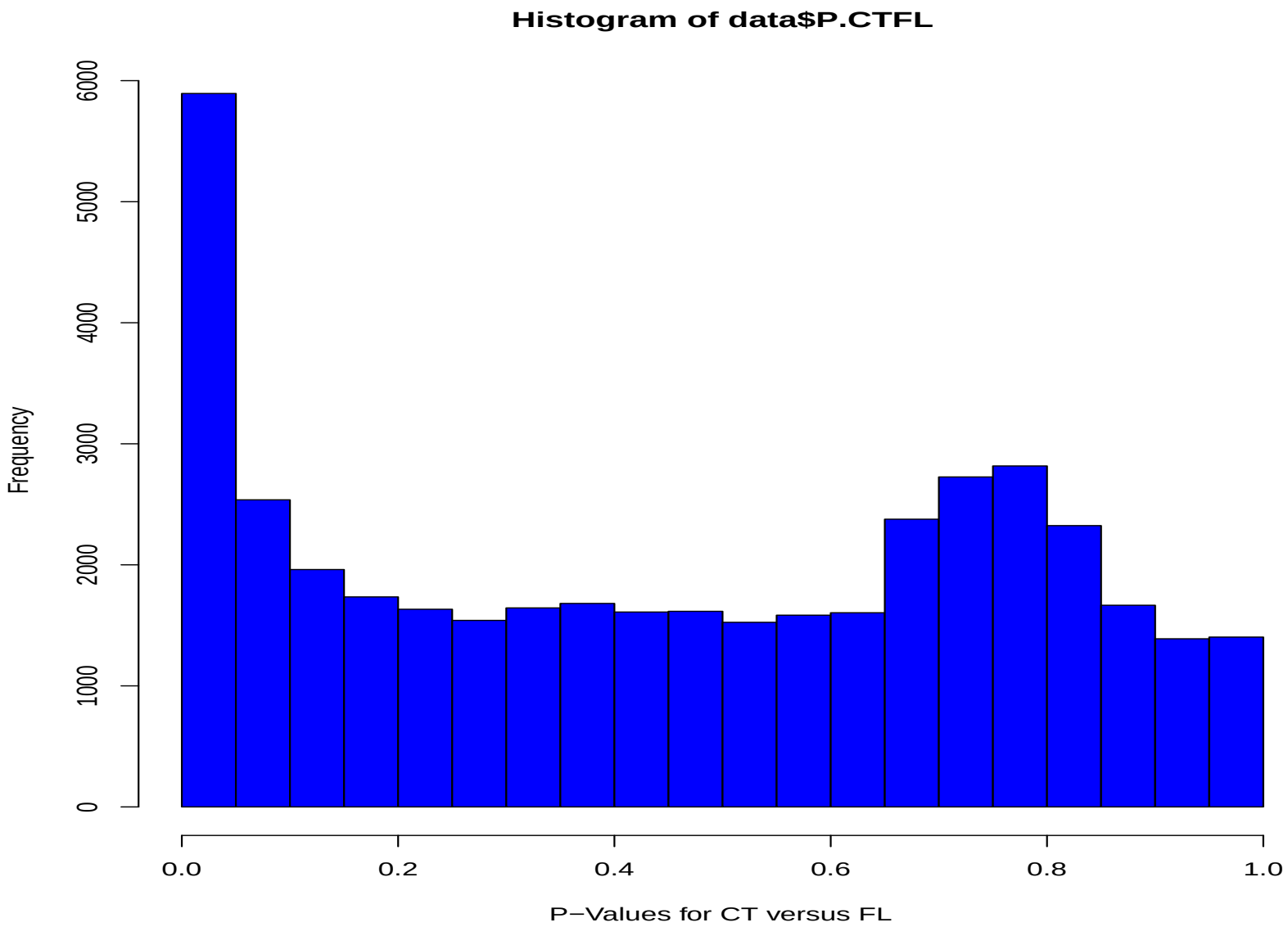
$$\delta_m^{BH}(X) = I\left\{P_m \leq P_{(K)}\right\}, \; m \in \mathcal{M}.$$

- Simple and easy-to-implement, but is it the BEST?

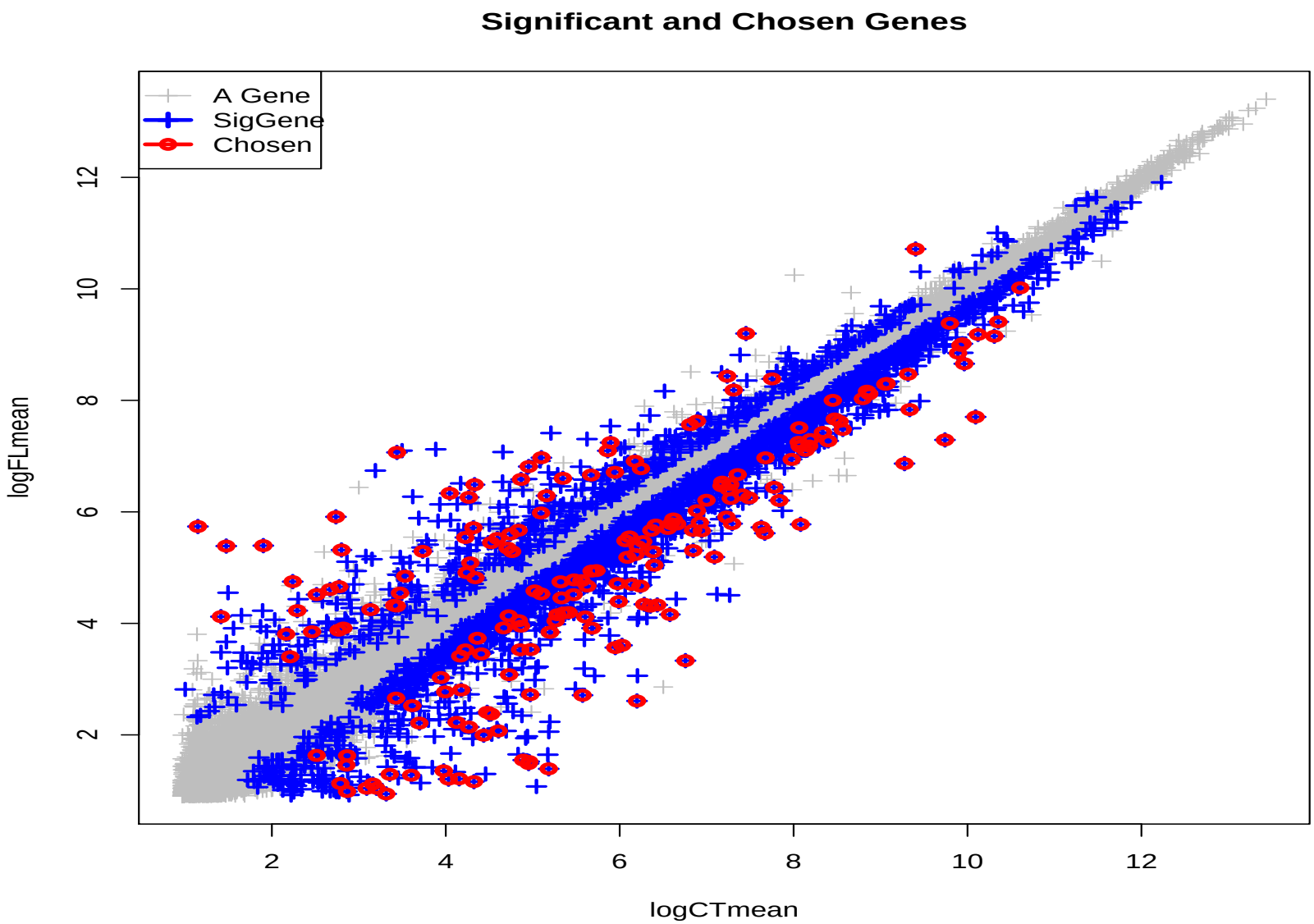
Applying BH Procedure to a Two-Group Microarray Data

- Agilent Technology microarray data set from M. Peña's lab. Jim Ryan of NOAA did the microarray analysis.
- $M = 41267$ genes.
- 2 groups, each group with 5 replicates.
- Applied t-test for each gene, using logged expression values. P-values obtained.
- Applied Benjamini-Hochberg Procedure with $q = .15$ to pick out the significant genes from the $M = 41267$ genes.
- Procedure picked out **2599** significant genes.
- Further analyzed the top (wrt to their p-values) 200 genes from these selected genes.
- Performed a cluster analysis on these 200 genes.

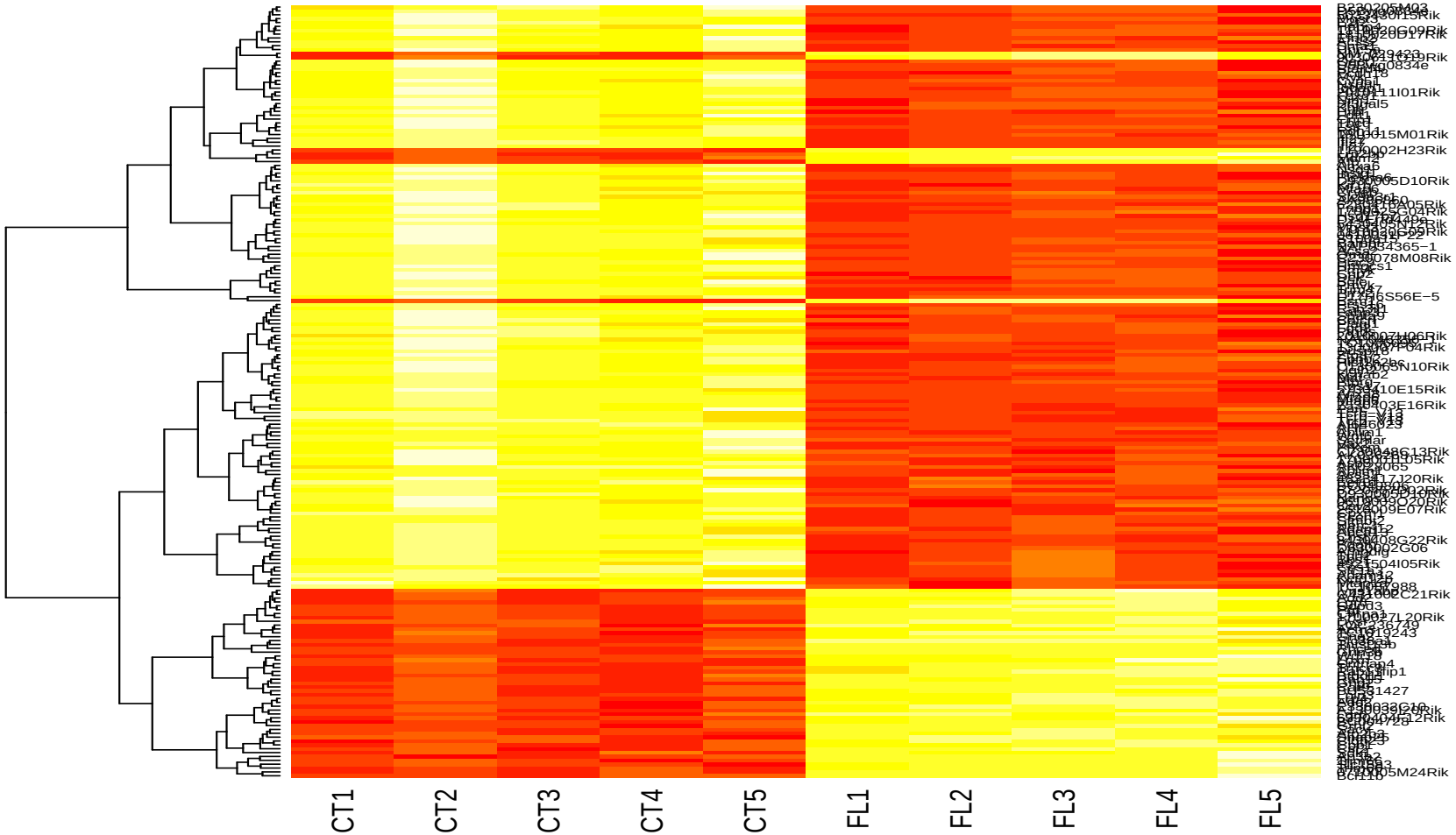
Histogram of the P-Values from the t-Tests



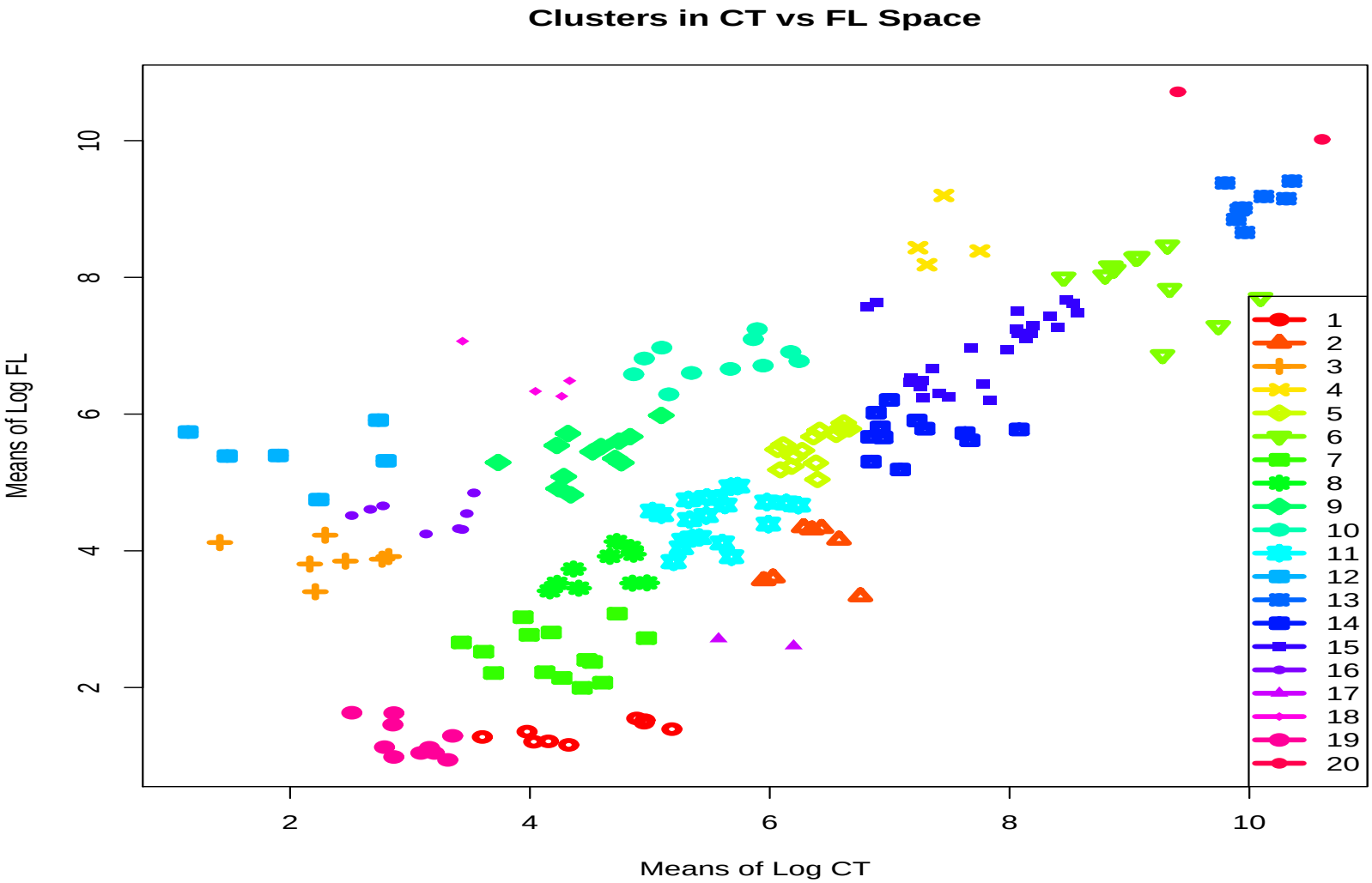
Scatterplot of the Pairwise Gene Means



Heatmap of the 200 Top Genes



Pictorial Depiction of Gene Clusters of Top 200 Genes



Can We Obtain a Better MDF than BH?

- **IDEA:** Given MDP $\Delta = (\Delta_m : m \in \mathcal{M})$, we find the optimal MDS $\mathbf{A}^* \equiv \mathbf{A}^*(\Delta) \in \mathfrak{S}$ achieving smallest MDR

$$R_2[(\Delta \circ \mathbf{A})(\alpha), P_1] = \frac{1}{M} \sum \{1 - \pi_m [A_m(\alpha)]\} .$$

- $\pi_m(\alpha) =$ POWER of $\delta_m(\alpha)$
- FWER-controlling MDF:

$$\delta^\dagger(q) = \delta^\dagger(q; \Delta, \mathbf{A}^*(\Delta))$$

- FDR-controlling MDF:

$$\delta^*(q) = \delta^*(q; \Delta, \mathbf{A}^*(\Delta))$$

- Use the best MDP Δ, e.g., MPs; UMPs; UMPUs; UMPIs.

Role of Power or ROC Functions

- P-value based procedures ignore differences in powers.
- Neyman and Pearson: power germane in search for optimality.
- Power of mth Test: $\pi_m(\alpha) = E_{P_{m1}}\{\delta_m(X;\alpha)\}$
- ROC Function for mth Decision Process Δ_m:

$$\alpha \mapsto \pi_m(\alpha)$$

- ROC functions in the missed discovery rate.
- Enables exploiting differences in the ROC functions.
- Why Power or ROC Differences? Different effect sizes, decision processes, or dispersion parameters.
- EXCHANGEABILITY: EXCEPTION rather than RULE!

Case with Simple Nulls and Simple Alternatives

- Neyman-Pearson Most Powerful Decision Process for each m.
- ROC Functions:

$$\alpha \mapsto \pi_m(\alpha)$$

- ROC functions are concave, continuous, and increasing.
- Assume that they are also twice-differentiable.

Theorem

Multiple decision size function $(\alpha \mapsto A_m(\alpha) : \ m \in \mathcal{M})$ *is optimal if it satisfies the* $M + 1$ *equilibrium conditions*

$$\forall m \in \mathcal{M} : \quad \pi'_m(A_m)(1 - A_m) = \lambda \quad \text{for some } \lambda \in \Re;$$

$$\sum_{\mathcal{M}} \log(1 - A_m) = \log(1 - \alpha).$$

Example: Optimal Multiple Decision Size Function

- $M = 2000$
- For each m: $X_m \sim N(\mu_m, \sigma = 1)$
- Multiple Decision Problem: To test

$$H_{m0} : \mu_m = 0 \quad \text{versus} \quad H_{m1} : \mu_m = \gamma_m.$$

- Effect Sizes: $\gamma_m \overset{IID}{\sim} |N(0, 3)|$
- For each m, Neyman-Pearson MP decision process.

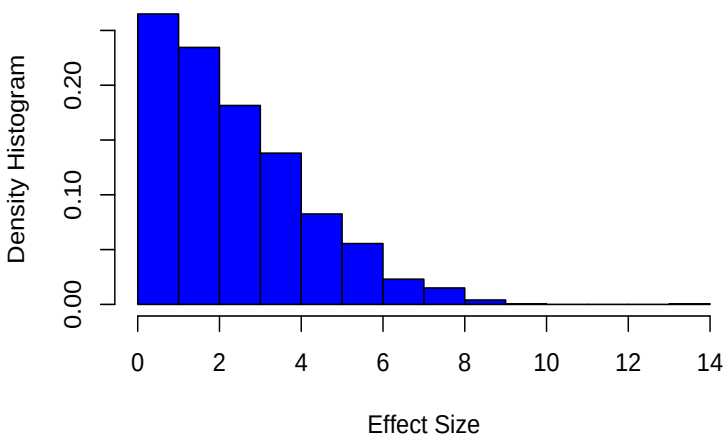
$$\Delta_m = (\delta_m(\alpha) : \alpha \in [0, 1])$$

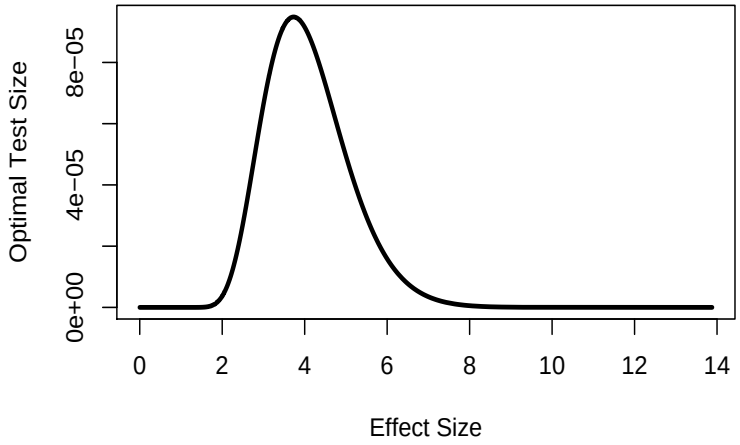
$$\delta_m(x_m; \alpha) = I\{x_m \geq \Phi^{-1}(1 - \alpha)\}$$

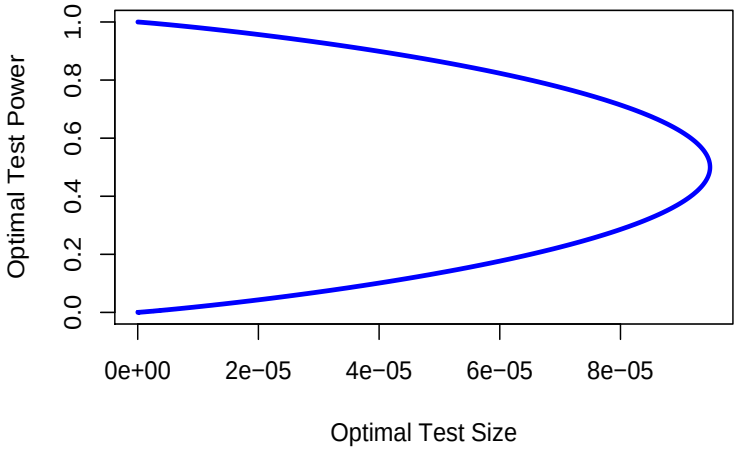
- Power or ROC Function for the mth NP MP Decision Process:

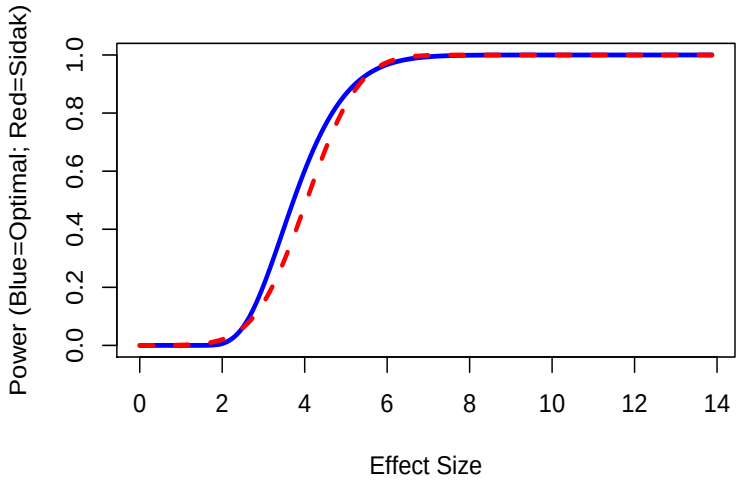
$$\alpha \mapsto \pi_m(\alpha) = 1 - \Phi\left[\Phi^{-1}(1 - \alpha) - \gamma_m\right]$$

Optimal Test Sizes vs Effect Sizes









Economic Aspect: A Size-Investing Strategy

- Do not invest your size on those where you will not make discoveries (small power) or those that you will certainly make discoveries (high power)!
- Rather, concentrate on those where it is a bit uncertain, since your differential gain in overall discovery rate would be greater!
- Some Wicked Consequences
  - Departmental Merit Systems.
  - Graduate Student Advising.

BH MDF versus $\delta^*(q)$: $q^* = .1$; $M = 20$; 1000 Reps

ν	p	δ_F^*-FDR	δ_F^*-MDR*	δ^{BH}-FDR	δ^{BH}-MDR*
1	0.1	8.03	70.80	8.43	72.64
1	0.2	7.55	79.64	8.77	81.99
1	0.4	6.05	77.47	6.65	80.30
2	0.1	7.70	54.42	8.43	55.80
2	0.2	7.39	56.32	7.59	57.31
2	0.4	6.47	47.82	6.21	49.38
4	0.1	9.14	8.62	9.48	10.30
4	0.2	7.80	7.34	6.97	9.20
4	0.4	6.15	3.58	5.65	5.53

BH MDF versus $\delta^*(q)$: $q^* = .1$; $M = 100$; 1000 Reps

ν	p	δ_F^*-FDR	δ_F^*-MDR*	δ^{BH}-FDR	δ^{BH}-MDR*
1	0.1	9.14	87.10	9.02	90.02
1	0.2	8.21	84.05	8.78	87.38
1	0.4	5.92	80.12	5.88	83.73
2	0.1	9.79	66.10	9.24	67.93
2	0.2	7.68	58.25	7.94	59.93
2	0.4	5.74	49.29	6.10	50.90
4	0.1	8.37	10.44	8.62	12.36
4	0.2	7.72	5.93	7.81	8.22
4	0.4	5.69	3.80	6.14	5.72

Potential Applications and Concluding Remarks

- Microarray data analysis: which genes are important?
- Systems analysis (Biological Pathways?): which components (subsystems of genes) are relevant?
- Variable selection: which predictor variables are important?
- For each gene, component, or predictor variable, apply a decision function to decide whether, say, independence or dependence holds with respect to the response variable.
- Test for Independence: Kendall's procedure, for example.
- Use MDFs $\delta^{\dagger}(q)$ or $\delta^{*}(q)$.
- Issues of determining effect sizes to determine power or ROC functions still need further studies.
- Comparison with other methods, such as those using regularization?

Acknowledgements

- Co-author Wensong Wu, my PhD student at USC Stat; joining FIU as Asst Prof this August.
- Co-author Josh Habiger, former PhD student at USC Stat; now Asst Prof at Oklahoma State.
- Marge Peña (Biology), Yu Zhang (Biology), and James Ryan (NOAA).
- Thanks to NSF, NIH, and EPA Grants which partially supported this work.
- Paper just appeared in the Feb 2011 issue of the *Annals of Statistics*.