# *P*-Values for Multiple Testing Procedures

Dissertation Defense

by: Joshua D. Habiger
Major Professor: Dr. Edsel Peña

Committee members:
Dr. Don Edwards
Dr. John Grego
Dr. Joshua Tebbs
Dr. Hongmei Zhang

Department of Statistics
University South Carolina

May 26, 2010

# Outline

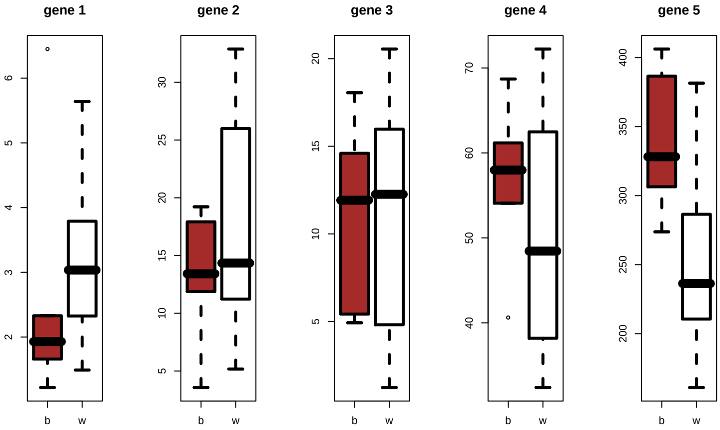
# Timmons et. al [2007] data

- $x_{mj}$ = brown fat cell expression measurement for gene $m$ on $j$th microarray
- $y_{mk}$ = white fat cell expression measurement for gene $m$ on $k$th microarray

| m | $x_{m1}$ | $x_{m2}$ | ... | $x_{m5}$ | $y_{m1}$ | $y_{m2}$ | ... | $y_{m8}$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 1.22 | 1.66 | ... | 2.33 | 5.64 | 1.79 | ... | 4.05 |
| 2 | 3.57 | 19.22 | ... | 11.89 | 5.17 | 29.49 | ... | 11.26 |
| $\vdots$ | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ |
| 12488 | 2.52 | 10.91 | ... | 22.67 | 10.70 | 7.35 | ... | 12.81 |

**Goal: compare brown fat cell measurements to white for each gene**

# Boxplots for 5 genes



# Model and Hypotheses

- Model:

$$X_{m1}, X_{m2}, \ldots, X_{m5} \overset{i.i.d.}{\sim} F_m(\cdot)$$

$$Y_{m1}, Y_{m2}, \ldots, Y_{m8} \overset{i.i.d.}{\sim} F_m(\cdot - \theta_m)$$

- Hypotheses:

$$H_{m0} : \theta_m = 0, F_m \in \mathcal{F}^{NORM} \text{ vs. } H_{m1} : \theta_m \neq 0, F_m \in \mathcal{F}^{NORM}$$

- $\theta_m = 0$ means gene$_m$ not differentially expressed
- $\theta_m \neq 0$ means gene$_m$ is differentially expressed

# Decision Functions

- Test statistics

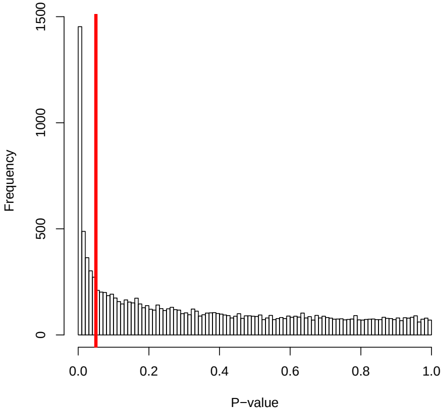
$$T(\boldsymbol{x}_m, \boldsymbol{y}_m) = \frac{\bar{x}_m - \bar{y}_m}{S_{pm}\sqrt{\frac{1}{5} + \frac{1}{8}}}$$

$$P_{T_m}(\boldsymbol{x}_m, \boldsymbol{y}_m) = 2[1 - \mathcal{T}_{11}(|T(\boldsymbol{x}_m, \boldsymbol{y}_m)|)]$$

- Decision Function

$$\delta_p(\boldsymbol{x}_m, \boldsymbol{y}_m; \alpha) = I(P_{T_m}(\boldsymbol{x}_m, \boldsymbol{y}_m) \leq \alpha)$$

# Testing with *P*-values




$\alpha = .05$ allows for 2879 "**DISCOVERIES**"!!!

# A closer look

- All $H_{m0}$ true => expect 624 False Discoveries!
- Consequences
  - Time (and grant money!?!) wasted
- **Solutions: Control global error rate**
  - $FWER = \Pr(\#FD \geq 1)$
  - $FDR = E\left[\frac{\#FD}{\max\{1, \#D\}}\right]$
  - . . . and many more

# P-Value based MTPs

- Let $\boldsymbol{P} = (P_1, P_2, ..., P_M)$ be $P$-values for testing $H_{10}, H_{20}, ..., H_{M0}$
- Define a $P$-value based MTP by

$$\boldsymbol{\delta} : [0,1]^M \to \{0,1\}^M$$

where $\boldsymbol{\delta}(\boldsymbol{P}) = (\delta_1(\boldsymbol{P}), \delta_2(\boldsymbol{P}), ..., \delta_M(\boldsymbol{P}))$

# Examples

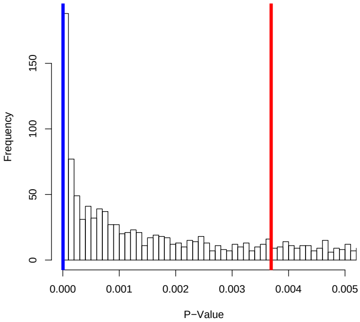
- Let $P_{(1)} \leq P_{(2)} \leq \ldots \leq P_{(M)}$ be ordered $P$-values
- Sequential Sidak (Duduit and van der Laan [2008])
  - $\delta_m(\boldsymbol{P}; \alpha) = I\left(P_m \leq \mathbf{1 - (1 - \alpha)^{\frac{1}{M-k+1}}}\right)$, where

$$k \equiv k(\boldsymbol{P}) = \max\left\{m : P_{(j)} \leq 1 - (1 - \alpha)^{\frac{1}{M-j+1}}, \forall j \leq m\right\}$$

- Benjamini and Hochberg (1995)
  - $\delta_m(\boldsymbol{P}; \alpha) = I(P_m \leq \boldsymbol{k\alpha/M})$, where

$$k \equiv k(\boldsymbol{P}) = \max\left\{j : P_{(j)} \leq \frac{j\alpha}{M}\right\}$$

# Application of MTPs: $\alpha = .05$




| Procedure | Discoveries | Error Rate Controlled? |
|---|---|---|
| BH | 922 | FDR |
| Seq Sid | 48 | FWER |

# Problems/Questions

- The fine print
  - Uniformity condition
  - Independence condition
- $T$-test $P$-values/ Nonparametric rank based $P$-values and the fine print
- Questions
  - **Could we define more *robust* $P$-value statistics satisfying the conditions? (relax normality assumption)**
  - **Could we define more *efficient* $P$-value statistics satisfying the conditions?**

# Randomized $P$-value Statistics for MTPs

# The Idea

- Use randomization to allow for nonparametric discrete $P$-values to be continuous
- Then $P$-values will satisfy uniformity condition and inherit robust properties

# Data description

- Assume $X \sim F$, let $\mathcal{F}$ be a model for $F$ and let $U \sim [0, 1]$ be independent of $X$.
- Let $\mathcal{F}_0$ be null sub-model for $F$ under $H_0$. i.e. $H_0 : F \in \mathcal{F}_0$ and $\mathcal{F}_1$ an alternative sub-model
- **Model example in this section**
  - $X_1, X_2, ..., X_{n_1} \overset{i.i.d.}{\sim} F(\cdot)$ and $Y_1, Y_2, ..., Y_{n_2} \overset{i.i.d.}{\sim} F(\cdot - \theta)$ with $\mathcal{F} = \{\text{all continuous d.f.s}\}$
  - $\mathcal{F}_0 = \{F \in \mathcal{F} : \theta \geq 0\}$ and $\mathcal{F}_1 = \{F \in \mathcal{F} : \theta < 0\}$

# Valid Decision Processes

- Decision function $\delta(x, u; \eta) \in \{0, 1\}$ where $\delta = 1(0)$ means reject(accept) $H_0$

- Allowing **size index** $\eta$ to vary, we form **decision process**

$$\Delta = \{\delta(X, U; \eta) : \eta \in [0, 1]\}$$

(also assume $t \mapsto \delta(x, u; t)$ nondecreasing and right cont. a.e. $[F] \forall F \in \mathcal{F}$)

**Definition**: *The decision process* $\Delta$ *is* *$\mathcal{F}_0$-size-valid* *if* $\sup_{F \in \mathcal{F}_0} E_{(F,U)}[\delta(X, U; \eta)] = \eta$ *for every* $\eta \in [0, 1]$.

# Uniform P-Value Statistics

$$P_\Delta(X, U) = \inf\{\eta \in [0, 1] : \delta(X, U; \eta) = 1\}$$

**Definition**: *The P-value statistic $P_\Delta(X, U)$ is $\mathcal{F}_0$-uniform if*

$$\sup_{F \in \mathcal{F}_0} Pr_{(F,U)}[P_\Delta(X, U) \le t] = t$$

*for every $t \in [0, 1]$.*

# Valid vs. Uniform

**Theorem**: $P_\Delta(X, U)$ ***is $\mathcal{F}_0$-uniform if and only if $\Delta$ is $\mathcal{F}_0$-size-valid.***

Proof: Show $[P_\Delta(X, U) \leq t] = [\delta(X, U; t) = 1]$ a.e $[F] \forall F \in \mathcal{F}$

- Use: We can usually define $\phi(x; \eta) \in [0, 1]$ so that $\sup_{F \in \mathcal{F}_0} E_F[\phi(x; \eta)] = \eta$.
- Then define $\delta(X, U; \eta) = I(U \leq \phi(X; \eta))$ and get $P_\Delta$

# Randomized Wilcoxon test function

- Randomized Wilcoxon test function:

$$\phi_{WR}(\boldsymbol{x},\boldsymbol{y};\eta) = \begin{cases} 1 & \text{if } W(\boldsymbol{x},\boldsymbol{y}) < k(\eta) \\ \gamma(\eta) & \text{if } W(\boldsymbol{x},\boldsymbol{y}) = k(\eta) \\ 0 & \text{if } W(\boldsymbol{x},\boldsymbol{y}) > k(\eta) \end{cases}$$

  - $k(\eta)$, $\gamma(\eta)$ are chosen s.t. $\sup_{F\in\mathcal{F}_0} E_F[\phi(\boldsymbol{x},\boldsymbol{y};\eta)] = \eta$

# Randomize Wilcoxon *P*-value

- Decision function:

$$\delta_{WR}(\boldsymbol{x},\boldsymbol{y},u;\eta) = I[u \leq \phi_W(\boldsymbol{x},\boldsymbol{y};\eta)]$$

$$= I[W(\boldsymbol{x},\boldsymbol{y}) < k(\eta)] + I[u \leq \gamma(\eta)]I[W(\boldsymbol{x},\boldsymbol{y}) = k(\eta)]$$

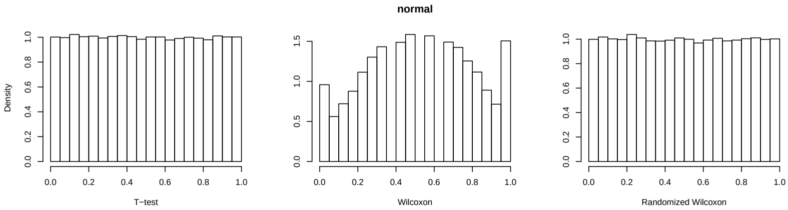
- *P*-value for $\Delta_{WR}$ is

$$P_{\Delta_{WR}}(\boldsymbol{x},\boldsymbol{y},u;\eta) = \inf\{\eta \in [0,1] : \delta_W(\boldsymbol{x},\boldsymbol{y},u;\eta) = 1\}$$

$$= \mathcal{W}_{n_1,n_2}[W(\boldsymbol{x},\boldsymbol{y}) - 1] + u w_{n_1,n_2}[W(\boldsymbol{x},\boldsymbol{y})]$$

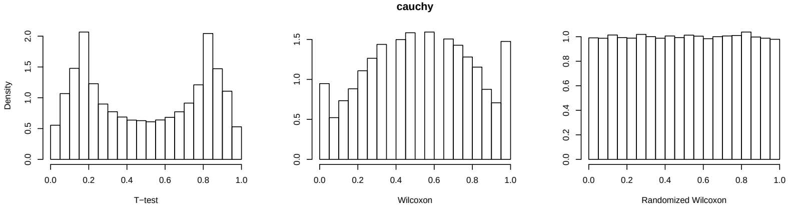
# *P*-value distribution: F=Normal

- 50,000 sets of $X_1, .., X_5, Y_1, ..., Y_5 \overset{i.i.d.}{\sim} F(\cdot)$ are generated and Wilcoxon, randomized Wilcoxon, and $T$-test $P$-values are computed.



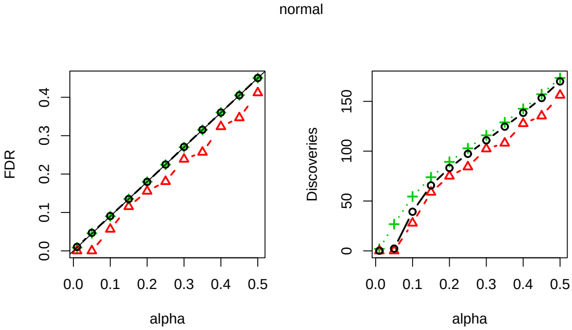
# *P*-value distribution: F=Cauchy

- 50,000 sets of $X_1, .., X_5, Y_1, ..., Y_5 \overset{i.i.d.}{\sim} F(\cdot)$ are generated and Wilcoxon, randomized Wilcoxon, and $T$-test $P$-values are computed.



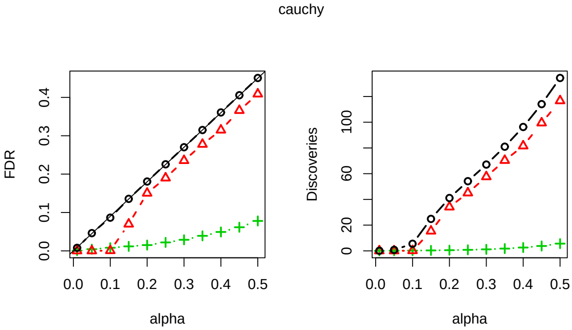
# BH Procedure for *F* = normal

- Same data except now $\theta_1 = \ldots = \theta_{900} = 0$ and $\theta_{901} = \ldots = \theta_{1000} = 2$
- BH procedure applied at $\alpha$ using *P*-values from T, Wilcoxon, and **Randomized Wilcoxon** tests.



# BH Procedure for $F$ = Cauchy

- Same data except now $\theta_1 = \ldots = \theta_{900} = 0$ and $\theta_{901} = \ldots = \theta_{1000} = 2$
- BH procedure applied at $\alpha$ using *P*-values from T, Wilcoxon, and **Randomized Wilcoxon** tests.



# Some Remarks

- **Randomized Wilcoxon $P$-value allows for valid MTPs so long as $F$ continuous**
  - $T$-test $P$-values only valid for normal model
  - Nonrandomized Wilcoxon $P$-values are never Uniform
- **What about efficiency?**

# Compound $P$-Value Statistics for MTPs

# The idea

- So far each $P$-value is *simple*
  - i.e. if $X_m$ is data for testing $H_{m0}$, then $P_{\Delta_m}(X_m)$
- What about *compound* $P$-value statistics?
  - i.e. for $\boldsymbol{X} = (X_1, X_2, ..., X_M)$, compute $P_{\Delta_m}(\boldsymbol{X})$
- How can we define *compound* $P$-value statistics and ensure ***uniform*** and ***independence*** conditions satisfied?

# Data Description

- Let $X \in \mathcal{X}$ and $X \sim F$.

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1N} \\ X_{21} & X_{22} & \cdots & X_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ X_{M1} & X_{M2} & \cdots & X_{MN} \end{bmatrix}$$

- For $\boldsymbol{A} \subseteq \mathcal{M} = \{1, 2, ..., M\}$, $\boldsymbol{B} \subseteq \mathcal{N} = \{1, 2, ..., N\}$ denote by

$$X[\boldsymbol{A}, \boldsymbol{B}] = (X_{mn} : m \in \boldsymbol{A}, n \in \boldsymbol{B})$$

- We write $X[m,] \equiv X[\{m\}, \mathcal{N}]$ to refer to a row $m$ and $X[,n]$ for column $n$

# Model

- Model: $X \sim F \in \mathcal{F}$
  - **Example for this section** - $N$ microarrays are i.i.d. according to an $M$-dimensional multivariate normal distribution

$$\mathcal{F} = \{F : F(x) = \prod_{n \in \mathcal{N}} G(x[, n]), G = MVN(\mu_M, \Sigma_{M \times M})\}$$

- Sub-models: $\mathcal{F}_{m0} \subseteq \mathcal{F}$ and $\mathcal{F}_{m1} \subseteq \mathcal{F}$
  - Ex. mean expression level for gene $m$ is 0. i.e.

$$\mathcal{F}_{m0} = \{F \in \mathcal{F} : \mu_m = 0\}$$

# Hypotheses

- Hypotheses

$$H_{m0} : F \in \mathcal{F}_{m0} \text{ vs. } H_{m1} : F \in \mathcal{F}_{m1}$$

- Global null model/hypothesis: For $\mathcal{M}_0 \subset \mathcal{M}$,

$$H_{\mathcal{M}0} : F \in \mathcal{F}_{\mathcal{M}_0} = \cap_{m \in \mathcal{M}_0} \mathcal{F}_{m0}$$

- Ex. mean expression level for genes 1 and 2 is 0

$$H_{\mathcal{M}0} : F \in \mathcal{F}_{\mathcal{M}0} = \{F \in \mathcal{F} : \mu_1 = \mu_2 = 0\}$$

# Size Valid Decision Processes

- To test $H_{m0} : F \in \mathcal{F}_{m0}$ with data $X \in \mathcal{X}$, we have compound decision function $\delta_m(X; \eta_m)$ with **compound decision process**

$$\Delta_m = \{\delta_m(X; \eta_m) : \eta_m \in [0, 1]\}$$

- We have Multiple Decision Process $\boldsymbol{\Delta} = (\Delta_m, m \in \mathcal{M})$

**Definition**: $\Delta_m$ is $\mathcal{F}_{m0}$-size valid if

$$\sup_{F \in \mathcal{F}_{m0}} E_F[\delta_m(X; \eta_m)] = \eta_m$$

# Uniform P-Value Statistics

**Definition**: *The **P-value statistic*** generated by $\Delta_m$ is defined via

$$P_{\Delta_m}(X) = \inf\{\eta_m \in [0,1] : \delta_m(X;\eta_m) = 1\}$$

**Definition**: $P_{\Delta_m}(X)$ is **$\mathcal{F}_{m0}$-uniform** if

$$\sup_{F \in \mathcal{F}_{m0}} \Pr_F(P_{\Delta_m}(X) \leq t) = t$$

for every $t \in [0,1]$.

**Theorem**: $P_{\Delta_m}(X)$ is $\mathcal{F}_{m0}$-uniform iff $\Delta_m$ is $\mathcal{F}_{m0}$-valid

# Independent $P$-value Statistic

**Definition**: $P_{\mathbf{\Delta}}(X) = (P_{\Delta_m}(X), m \in \mathcal{M})$ is $\mathcal{F}_{\mathcal{M}0}$-**independent** if for $\boldsymbol{t} \in [0,1]^M$ and $F \in \mathcal{F}_{\mathcal{M}0}$,

$$\Pr_F\left(\bigcap_{m \in \mathcal{M}} [P_{\Delta_m}(X) \le t_m]\right) = \Pr_F\left(\bigcap_{m \in \mathcal{M}_1} [P_{\Delta_m}(X) \le t_m]\right) \prod_{m \in \mathcal{M}_0} \Pr_F(P_{\Delta_m}(X) \le t_m)$$

- Question: When will an MDP generate $\mathcal{F}_{\mathcal{M}0}$-independent $P$-value statistics?

# Independence Theorem

**Definition:** $\boldsymbol{\Delta}$ is $\mathcal{F}_{\mathcal{M}0}$-independent if for $\boldsymbol{d} \in \{0,1\}^M$, $\boldsymbol{\eta} \in [0,1]^M$, and every $F \in \mathcal{F}_{\mathcal{M}0}$,

$$\Pr_F\left(\bigcap_{m\in\mathcal{M}}[\delta_m(X;\eta_m)=d_m]\right)$$
$$=\Pr_F\left(\bigcap_{m\in\mathcal{M}_1}[\delta_m(X;\eta_m)=d_m]\right)\prod_{m\in\mathcal{M}_0}\Pr_F\left(\delta_m(X;\eta_m)=d_m\right)$$

**Theorem:** ***$P_{\boldsymbol{\Delta}}(X)$ is $\mathcal{F}_{\mathcal{M}0}$-independent iff $\boldsymbol{\Delta}$ is $\mathcal{F}_{\mathcal{M}0}$ independent***

Proof follows by showing $I(P_{\Delta_m}(X)\le t_m)=\delta_m(X;t_m)$ a.e. [F] $\forall F$

# How can we define $\mathcal{F}_{\mathcal{M}0}$-independent and valid compound decision processes?

# The Idea

- Consider testing

$$H_{m0} : X_m \sim N(0,1) \text{ vs. } H_{1m} : X_m \sim N(\mu_m, 1), \mu_m \neq 0$$

- The standard decision function is defined

$$\delta_m(X_m; .05) = I(X_m \leq -1.96) + I(X_m \geq 1.96)$$

- Why not $\delta_m(X_m; .05) = I(X_m \geq 1.645)$?
  - If $\mu_m > 0$, NP test!
- We will split $X = (X[, T], X[, \bar{T}])$ and use $X[, T]$ to estimate $\mu_m$ and $X[m, \bar{T}]$ to test $H_{m0}$

# The Decision Function

- Consider compound decision functions $\delta_m : \mathcal{X} \rightarrow \{0, 1\}$ defined

$$\delta_m(X[, T], X[m, \bar{T}]; \eta)$$

where $T \subset \mathcal{N}$ and $\bar{T} = \mathcal{N} \setminus T$

- Ex. $T = \{1, 2\}$, and $\delta_1(X[, T], X[1, \bar{T}]; \eta)$

$$\begin{array}{cc|ccc} X_{11} & X_{12} & X_{13} & \cdots & X_{1N} \\ X_{21} & X_{22} & X_{23} & \cdots & X_{2N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_{M1} & X_{M2} & X_{M3} & \cdots & X_{MN} \end{array}$$

# The Example

- Suppose $X[,n] \overset{i.i.d}{\sim} F \in \mathcal{F} = \left\{MVN\left(\frac{1}{N}\boldsymbol{\mu}, \frac{1}{N}I\right)\right\}$
- $H_{m0} : \mu_m = 0, F \in \mathcal{F}$ vs. $H_{m1} : \mu_m \neq 0, F \in \mathcal{F}$
- Considering $M$ sufficient statistics for training data -$\sum_{n \in T} X[,n]$ - and test data - $\sum_{n \in \bar{T}} X[,n]$

$$
\begin{aligned}
\boldsymbol{X} &\equiv \sum_{n \in \mathcal{N}} X[,n] \\
&= \sum_{n \in T} X[,n] + \sum_{n \in \bar{T}} X[,n] \\
&\equiv \boldsymbol{Y} + \boldsymbol{Z}
\end{aligned}
$$

# The Example Cont.

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_M \end{bmatrix} \overset{\text{split}}{\longrightarrow} \begin{bmatrix} Y_1 & Z_1 \\ Y_2 & Z_2 \\ \vdots & \vdots \\ Y_M & Z_M \end{bmatrix}$$

- $\boldsymbol{X} \sim MVN(\boldsymbol{\mu}, I)$.
- $\boldsymbol{Y} \sim MVN(\lambda^2\boldsymbol{\mu}, \lambda^2 I)$ and $\boldsymbol{Z} \sim MVN((1-\lambda^2)\boldsymbol{\mu}, (1-\lambda^2)I)$
- $\lambda^2 = \frac{|T|}{N}$ is proportion of training data

# Simple and Compound Decision Functions

1. $\delta_m^{(1)}(Y_m, Z_m; \eta) \equiv \delta_m^{(1)}(X_m; \eta)$ - Simple
2. $\delta_m^{(2)}(\boldsymbol{Y}, Z_m; \eta)$ - Compound

- Ex. $\delta_1^{(1)}(Y_1, Z_1; \eta)$ vs. $\delta_1^{(2)}(\boldsymbol{Y}, Z_1; \eta)$

$$\begin{bmatrix} \boldsymbol{Y}_1 & \boldsymbol{Z}_1 \\ Y_2 & Z_2 \\ \vdots & \vdots \\ Y_M & Z_M \end{bmatrix} \text{ vs. } \begin{bmatrix} \boldsymbol{Y}_1 & \boldsymbol{Z}_1 \\ \boldsymbol{Y}_2 & Z_2 \\ \vdots & \vdots \\ \boldsymbol{Y}_M & Z_M \end{bmatrix}$$

# Evaluating MDPs

- $\boldsymbol{\Delta}^{(1)}$ and $\boldsymbol{\Delta}^{(2)}$ should each be $\mathcal{F}_{\mathcal{M}0}$-size valid (and independent)
- Want to maximize power

$$\beta^{(1)}(\boldsymbol{\mu}, \eta) = \sum_{m \in \mathcal{M}_1} \frac{\beta_m^{(1)}(\mu_m, \eta)}{M_1} \equiv \sum_{m \in \mathcal{M}_1} \frac{E_{\mu_m}[\delta_m^{(1)}(X_m; \eta)]}{M_1}$$

$$\beta^{(2)}(\boldsymbol{\mu}, \lambda^2, \eta) = \sum_{m \in \mathcal{M}_1} \frac{\beta_m^{(2)}(\boldsymbol{\mu}, \lambda^2, \eta)}{M_1} \equiv \sum_{m \in \mathcal{M}_1} \frac{E_{\boldsymbol{\mu}}[\delta_m^{(2)}(\mathbf{Y}, Z_m; \eta)]}{M_1}$$

# Gold Standard Decision Function

$$\delta_m^{(1)}(X_m;\eta) = I(X_m \le l_m(\eta)) + I(X_m \ge u_m(\eta))$$

- If $l_m(\eta) = \Phi^{-1}(\eta/2)$ and $u_m(\eta) = \Phi^{-1}(1-\eta/2)$, then

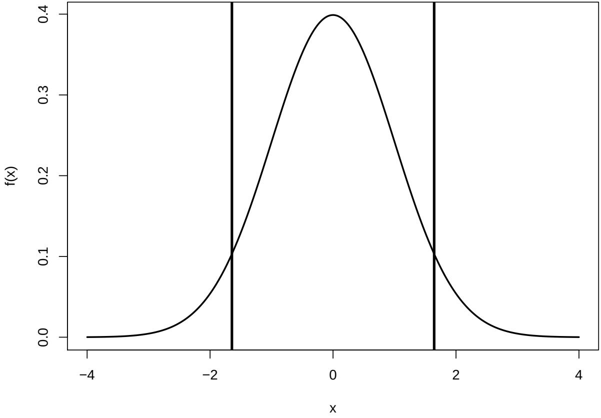
$$E_{\mu_m=0}[\delta_m^{(1)}(X_m;\eta)] = \eta$$

- Ex. $l(.1) = -1.645$ and $u(.1) = 1.645$

# Oracle Decision Function

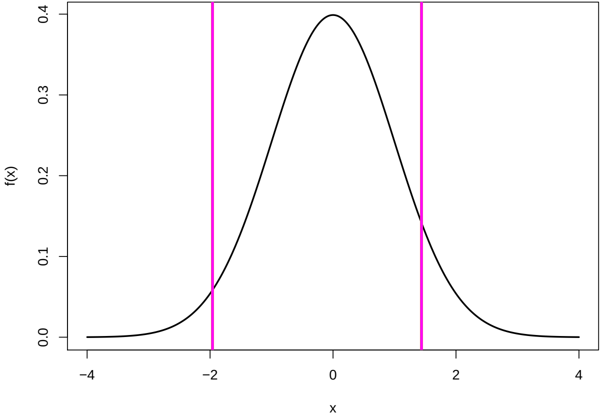
- How would an *Oracle*, who knew $\mu_m$, choose $l_m(\eta)$ and $u_m(\eta)$?
  1. Constraint: Need to choose $l_m(\eta)$ and $u_m(\eta)$ s.t.
  $$E_{\mu_m=0}[\delta_m^{(1)}(X_m;\eta)] = \eta$$
  2. Maximize $\beta_m(\mu_m, \eta)$

# Constraint



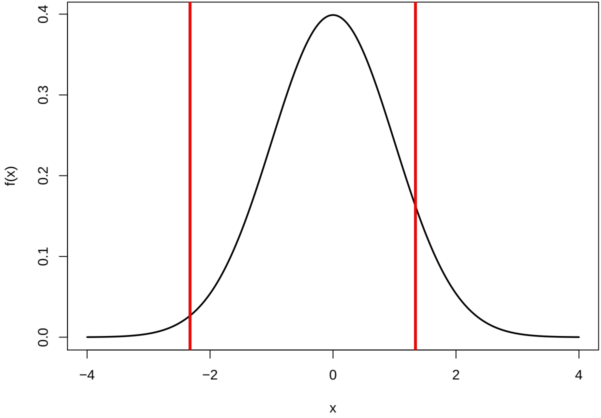

- Tail area is .1

# Constraint



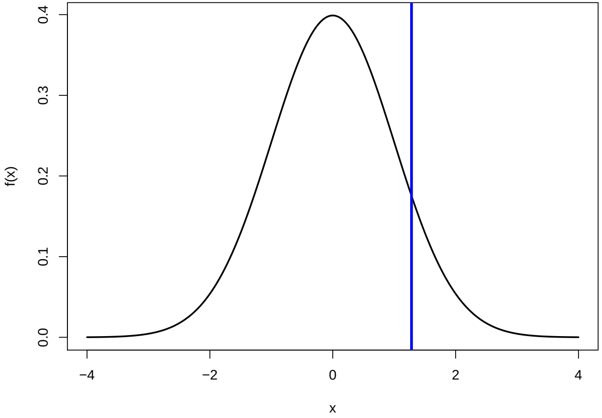

- Tail area is .1

# Constraint




- Tail area is .1

# Constraint




- Tail area is .1

# Oracle Cutoffs

- Constraint corresponds to $h_m \in [0,1]$ for
  - $l(\eta, h_m) = \Phi^{-1}(h_m \eta)$
  - $u(\eta, h_m) = \Phi^{-1}(1 - [1 - h_m]\eta)$
- For

$$\delta_m^{(1)}(X_m; \eta, h_m) = I(X_m \leq l(\eta, h_m)) + I(X_m \geq u(\eta, h_m)),$$

$$\beta_m^{(1)}(\mu_m, \eta, h_m) = \Phi(\mu_m - l(\eta, h_m)) + 1 - \Phi(\mu_m - u(\eta, h_m))$$

is maximized by choosing

$$h_m(\mu_m) = \begin{cases} 1 & \text{if } \mu_m < 0 \quad \text{(lower tailed test)} \\ 0 & \text{if } \mu_m > 0 \quad \text{(upper tailed test)} \end{cases}$$

# Estimating the Oracle Cutoffs

**Theorem**: *Suppose that $((Y_m, Z_m); m \in \mathcal{M}_0)$ are independent and also independent of $((Y_m, Z_m); m \in \mathcal{M}_1)$. If for every $m \in \mathcal{M}_0$,*

$$E_F[\delta_m(\mathbf{Y}, Z_m; \eta_m)|\mathbf{Y}] = \eta_m$$

*for every $F \in \mathcal{F}_{m0}$, then $\mathbf{\Delta}$ is $\mathcal{F}_{\mathcal{M}0}$-size valid and hence, $P_{\mathbf{\Delta}}(\mathbf{Y}, \mathbf{Z})$ is $\mathcal{F}_{\mathcal{M}0}$-uniform and independent.*

Intuition: $\delta_m$ only depends on $Z_m$ under $H_{m0}$ and $Z_m$s are independent! Also, $\delta_m$ is size valid.

# Estimating the Oracle Cutoffs Cont.

- Estimate $h_m(\mu_m) = I(\mu_m < 0)$ with $h_m(\boldsymbol{Y}) \in [0,1]$

$$\delta_m^{(2)}(\boldsymbol{Y}, Z_m; \eta) =$$
$$I\left(\frac{Z_m}{\sqrt{1-\lambda^2}} \leq l(h_m(\boldsymbol{Y}), \eta)\right) + I\left(\frac{Z_m}{\sqrt{1-\lambda^2}} \geq u_m(h_m(\boldsymbol{Y}), \eta)\right)$$

**Corollary:** $\boldsymbol{\Delta}^{(2)}$ ***is*** $\mathcal{F}_{\mathcal{M}0}$***-size-valid and independent for any*** $\mathcal{M}_0$***. Hence*** $P_{\boldsymbol{\Delta}^{(2)}}$ ***is*** $\mathcal{F}_{\mathcal{M}0}$***-uniform and independent for any*** $\mathcal{M}_0$***.***

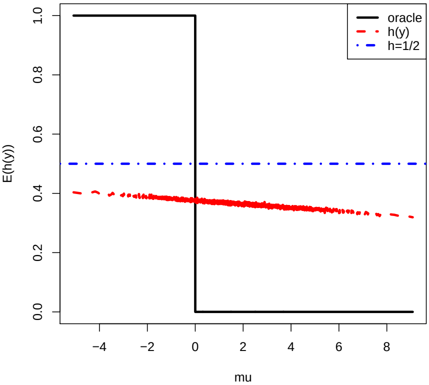
- Intuition: The tail area is $\eta$ for any $h_m(\boldsymbol{Y}) \in [0,1]$

# Estimating the Oracle Cutoffs Cont.

- How will we estimate $h_m(\mu_m) = I(\mu_m < 0)$ with $h_m(\boldsymbol{Y})$?
- **Route:** Use Empirical Bayes methods to develop shrinkage estimators
  1. Specify prior for $\mu_m$:
  $$G(\mu_m; \theta, \tau) = \Phi\left(\frac{\mu_m - \theta}{\tau}\right)$$
  2. Compute $h_m(Y_m, \theta, \tau) = \Pr(\mu_m < 0; Y_m, \theta, \tau)$
  3. Plug in MOM estimates $\hat{\theta}(\boldsymbol{Y})$ and $\hat{\tau}(\boldsymbol{Y})$

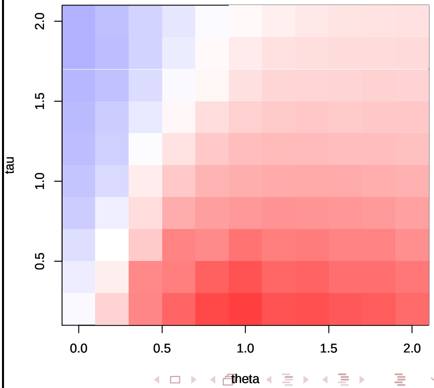
# Performance: Train Prop. = .001

$\mu_m$ vs. $h_m(\boldsymbol{Y}) : \theta = \tau = 2$



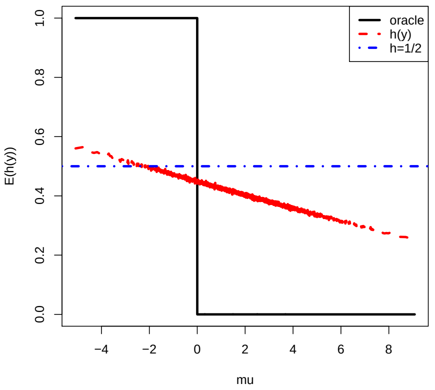

$$\frac{\beta^{(2)}(\boldsymbol{\mu}, \lambda^2, .05/5000)}{\beta^{(1)}(\boldsymbol{\mu}, .05/5000)}$$

.4 .7 .9 1.1 1.3 1.6



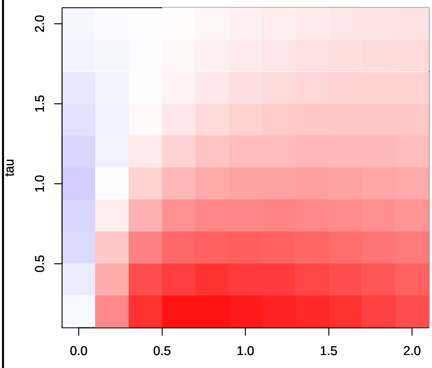
# Performance: Train Prop. = .01

$\mu_m$ vs. $h_m(\boldsymbol{Y}) : \theta = \tau = 2$




$$\frac{\beta^{(2)}(\boldsymbol{\mu}, \lambda^2, .05/5000)}{\beta^{(1)}(\boldsymbol{\mu}, .05/5000)}$$

.4 .7 .9 1.1 1.3 1.6

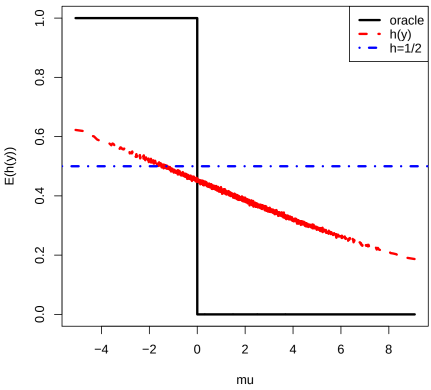


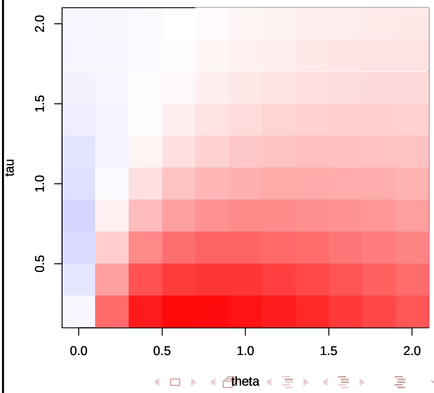
# Performance: Train Prop. = .02

$\mu_m$ vs. $h_m(\boldsymbol{Y}) : \theta = \tau = 2$

$$\frac{\beta^{(2)}(\boldsymbol{\mu}, \lambda^2, .05/5000)}{\beta^{(1)}(\boldsymbol{\mu}, .05/5000)}$$

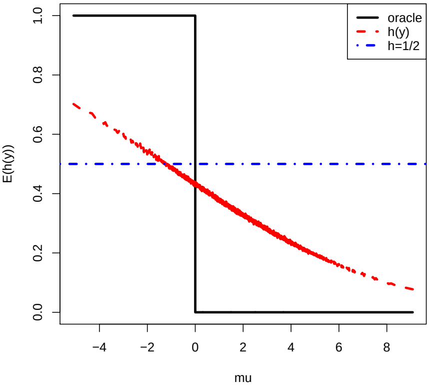
.4 .7 .9 1.1 1.3 1.6



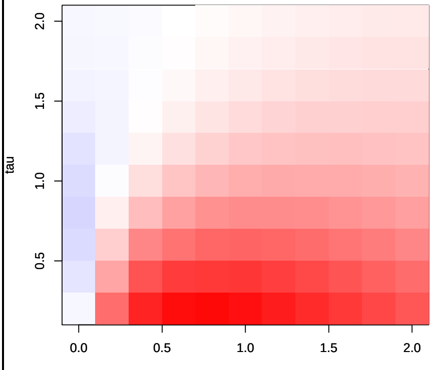
# Performance: Train Prop. = .05

$\mu_m$ vs. $h_m(\boldsymbol{Y}) : \theta = \tau = 2$



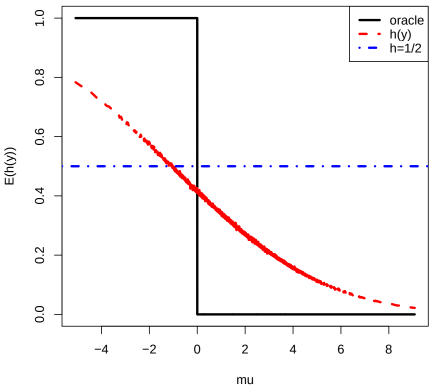

$$\frac{\beta^{(2)}(\boldsymbol{\mu}, \lambda^2, .05/5000)}{\beta^{(1)}(\boldsymbol{\mu}, .05/5000)}$$

.4 .7 .9 1.1 1.3 1.6



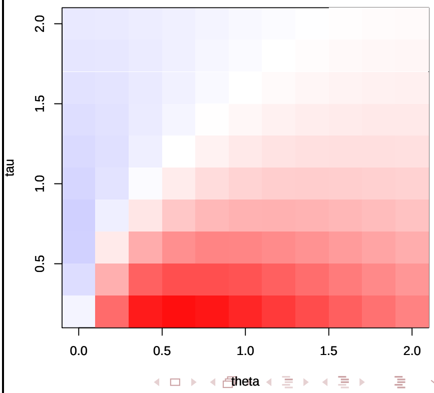
# Performance: Train Prop. = .1

$\mu_m$ vs. $h_m(\boldsymbol{Y})$ : $\theta = \tau = 2$



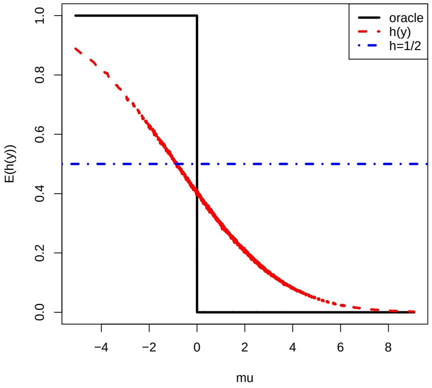

$$\frac{\beta^{(2)}(\boldsymbol{\mu},\lambda^2,.05/5000)}{\beta^{(1)}(\boldsymbol{\mu},.05/5000)}$$

.4 .7 .9 1.1 1.3 1.6



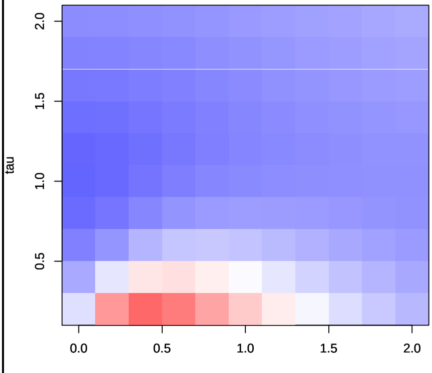
# Performance: Train Prop. = .2

$\mu_m$ vs. $h_m(\boldsymbol{Y}) : \theta = \tau = 2$




$\dfrac{\beta^{(2)}(\boldsymbol{\mu},\lambda^2,.05/5000)}{\beta^{(1)}(\boldsymbol{\mu},.05/5000)}$

.4 .7 .9 1.1 1.3 1.6



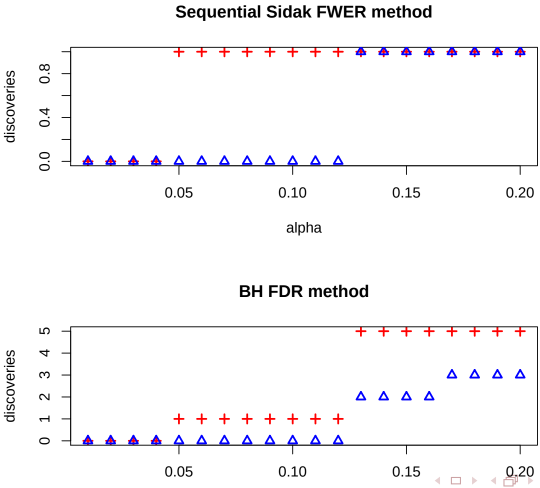
# Splitting the sample

- Choice of $\lambda^2$
  - Larger $\lambda^2$ allows for better estimates of $h_m(\mu_m)$ but we have less test data
  - Smaller $\lambda^2$ yields worse estimates of $h_m(\mu_m)$ but we have more test data
  - We should choose $.01 < \lambda^2 < .05$
- Caveats
  - If average signal is $\theta = 0$, simple $\delta_m$ is more powerful
  - Loss in power when $\theta = 0$ is small relative to gain in power when $\theta \neq 0$.

# Prostate Cancer Data

| | control group | | | | cancer group | | | |
|---|---|---|---|---|---|---|---|---|
| | $x[,1]$ | $x[,2]$ | ... | $x[,50]$ | $x[,51]$ | $x[,52]$ | ... | $x[,102]$ |
| $x[1,]$ | -.931 | -.840 | ... | 3.81 | -1.12 | 1.01 | ... | -.001 |
| $x[2,]$ | -1.07 | -.880 | ... | -.477 | -.571 | -.811 | ... | -.836 |
| $\vdots$ | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ |
| $x[6033,]$ | -.754 | -.708 | ... | -.011 | .457 | .578 | ... | -.162 |

# Application to BH and Sidak



# Some Remarks

- We allow for MTPs to depend upon ***compound*** $P$-values so the behave in a more efficient manner
- Sample splitting approach allows for MTPs to be valid, contrary to the double dipping approach in Sun and Cai [2007], Efron [2001,2004,2007,...]
- May be possible to use training data more efficiently. Choice of $\lambda^2$ will depend on $h_m(Y)$

# What have we done?

- There are many $P$-value based multiple testing procedures for controlling many different error rates
- Stochastic process approach allows for (possibly compound) $P$-values satisfying conditions allowing for valid MTPs
    - Robust
    - Efficient
- Methods can be broadly used to improve any $P$-Value based MTP

# Future Work

- Can use sample splitting approach in other Oracle procedures - Peña et. al.
- Investigate other $h_m(Y)$
- Bayesian rather than Empirical Bayesian approach (specify $G(\mu_m)$ and study robustness)
- Simultaneous conf. intervals. For $\theta$ a parameter, the $1-\alpha$ interval is $A(X) = \{\theta \in \Theta : P(X|\theta) \geq \alpha\}$

# Another Procedure

- Use all data to satisfy uniformity condition?
- Some Error rates - *mFDR*, *EFP* - don't require independence

Let $h_1(\boldsymbol{X}) = g(X_2, X_3, ..., X_M)$, $h_2(\boldsymbol{X}) = g(X_1, X_3, ..., X_M)$, ...

$$\delta_1(\boldsymbol{X};\eta_m) = I\left(X_1 \leq \Phi^{-1}(h_1(\boldsymbol{X})\eta)\right) + I\left(X_1 \geq \Phi^{-1}(1 - [1 - h_1(\boldsymbol{X})]\eta_m)\right)$$

$$\delta_2(\boldsymbol{X};\eta_m) = I\left(X_2 \leq \Phi^{-1}(h_2(\boldsymbol{X})\eta)\right) + I\left(X_2 \geq \Phi^{-1}(1 - [1 - h_2(\boldsymbol{X})]\eta_m)\right)$$

$$\vdots$$

# Another Procedure

- We chose $|T| = |T_1 \bigcup T_2| = 4$ in example
- Consider (unbiased) power estimate

$$\hat{\beta}_m = \frac{1}{\binom{50}{2}\binom{52}{2}} \sum_{T:|T_1|=|T_2|=2} \delta_m(x[,T], x[m,\bar{T}])$$

- Can simply report $\hat{\beta}_m$s or even define $\delta_m^* = I(U_m \leq \hat{\beta}_m)$ if you dare.

# Acknowledgements


- EPA
- NSF
- NIH
- Advisor
- Committee Members
- Department
- Kate