Compound p-Values for Multiple Testing Procedures

Presentation by:
Joshua D. Habiger (Oklahoma State University)

Joint work with:
Edsel Peña (University of South Carolina)

June 24, 2011

Prostate Cancer Data: Efron 2009

- N=102 microarrays and M=6033 gene expression measurements from each

m	control group				cancer group			
	1	2	...	50	1	2	...	52
1	-.931	-.840	...	3.81	-1.12	1.01	...	-.001
2	-1.07	-.880	...	-.477	-.571	-.811	...	-.836
⋮	⋮	⋮	⋱	⋮	⋮	⋮	⋱	⋮
6033	-.754	-.708	...	-.011	.457	.578	...	-.162

- **Which genes are related to cancer (differentially expressed)?**
- Need to test 6033 null hypotheses!

Strategy

DATA

$\Downarrow$

p-Values (6033 of them)

$\Downarrow$

Multiple Testing Procedure

$\Downarrow$

Decisions (6033 of them)

Strategy

DATA

$\Downarrow$

p-Values (6033 of them)

$\Downarrow$

Multiple Testing Procedure

$\Downarrow$

Decisions (6033 of them)

Properties of *p*-Values

Multiple testing procedures **valid** if *p*-values from **null hypotheses** are

1. **uniformly distributed** and
2. **independent**

Decision Function and p-Value

How to get a p-value for testing H_0 with data X

In words:
The p-value is " the smallest **size** allowing for the null to be rejected with the observed **data** x"

Mathematically:

1. Define decision function $\delta(x;\eta) \in \{0,1\}$
   - η is size and X is data
2. Definition: p-Value is $P(x) = \inf\{\eta : \delta(x;\eta) = 1\}$

We can work with $\delta(X;\eta)$ and use the previous definition to get a $P(X)$.

See Habiger and Pena (2011), JNS or Pena, Habiger, Wu (2011), AoS for details

Notation and Simple Model

$$\begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1N} \\ X_{21} & X_{22} & \cdots & X_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ X_{M1} & X_{M2} & .. & X_{MN} \end{bmatrix} \rightarrow \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_M \end{bmatrix}$$

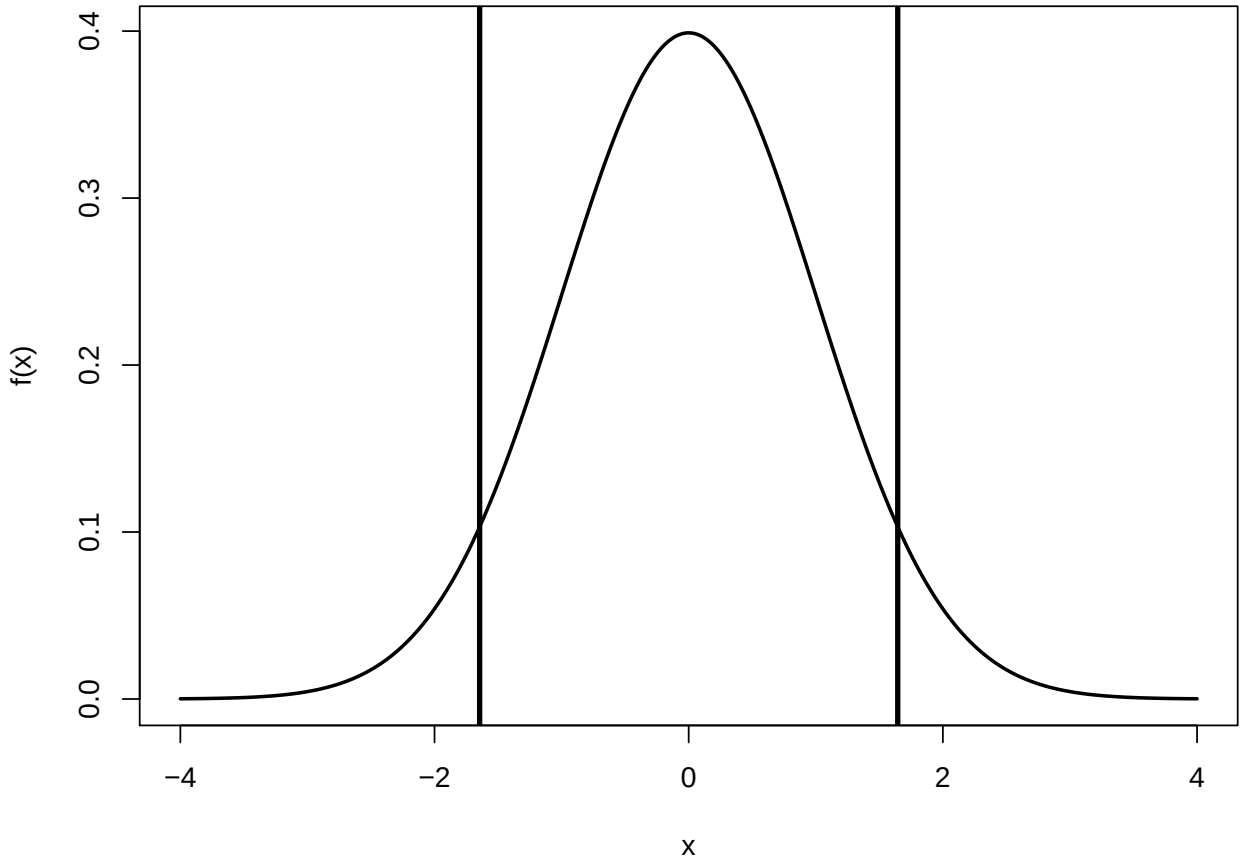
- $\rightarrow$ means test stat applied to each row and transformed to get $H_{0m} : X_m \sim N(0, 1)$
  - Ex. Compute $T_m = T(X_{m1}, X_{m2}, ..., X_{mN})$ and take $X_m = \Phi^{-1}(F_T(T_m))$

Basic Example

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

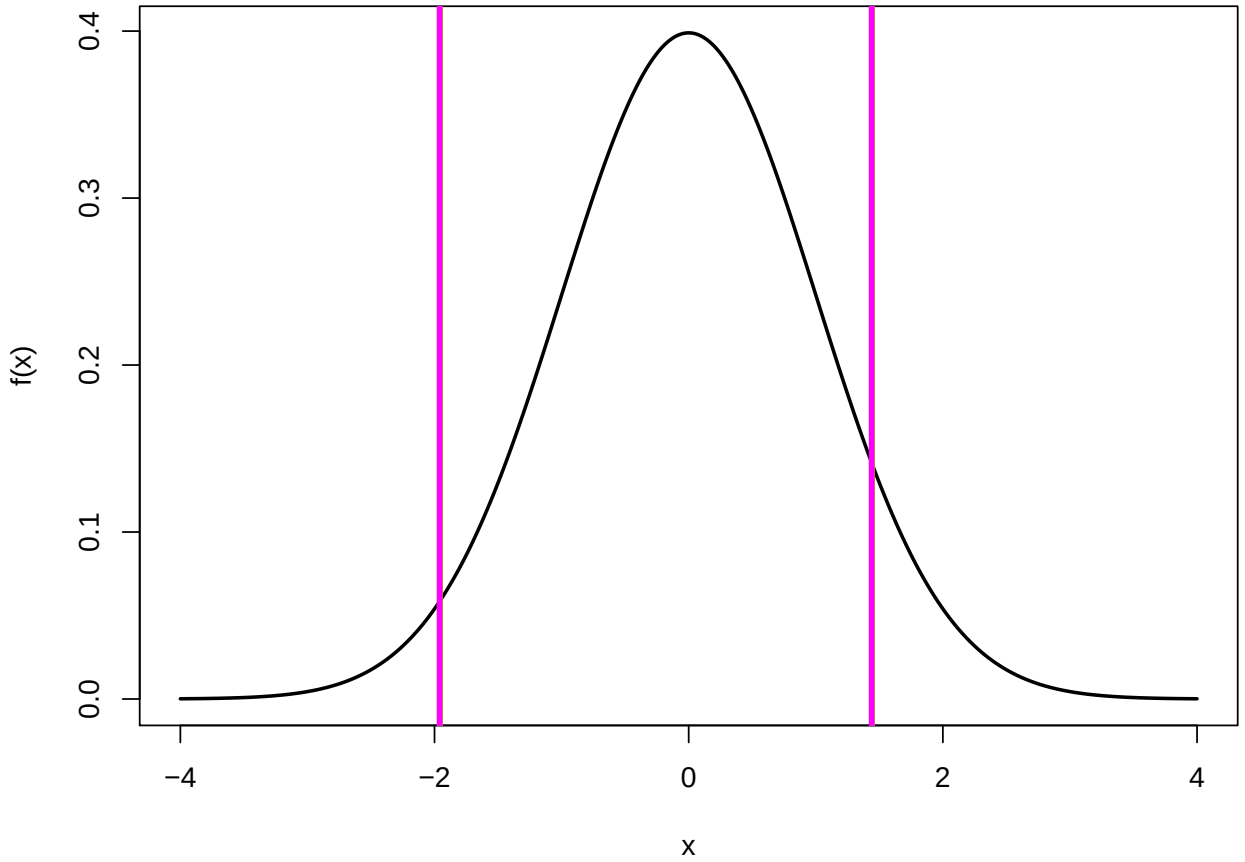
$$\delta_m(x_m; \eta_m) = \begin{cases} 1 & \text{if} \quad x_m > \Phi^{-1}(1 - \eta_m/2) \\ 1 & \text{if} \quad x_m < \Phi^{-1}(\eta_m/2) \\ 0 & \quad \textit{otherwise} \end{cases}$$

Example



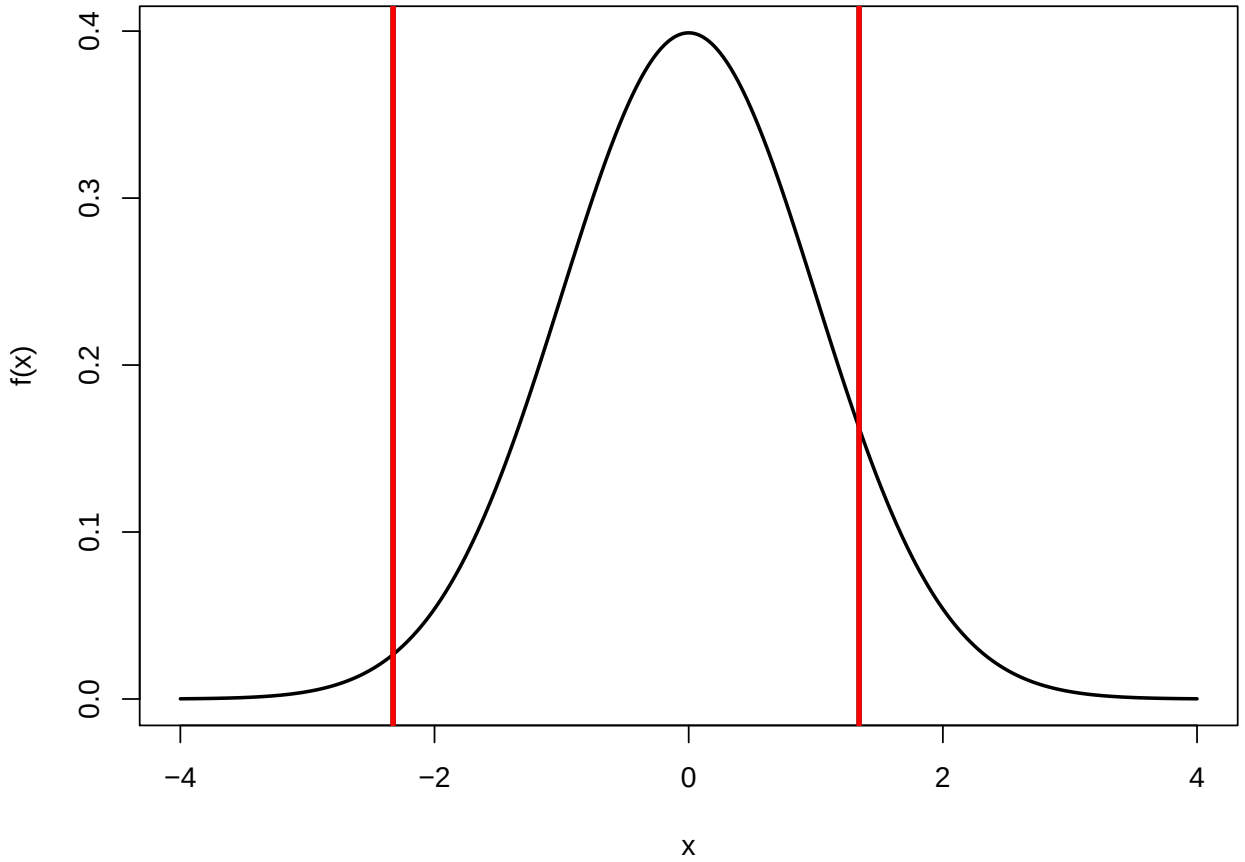
- Upper cutoff: $\Phi^{-1}(1 - .1/2) = 1.645$
- Lower Cutoff: $\Phi^{-1}(.1/2) = -1.645$
- **Tail area is $\eta_m = .1$!**

Example



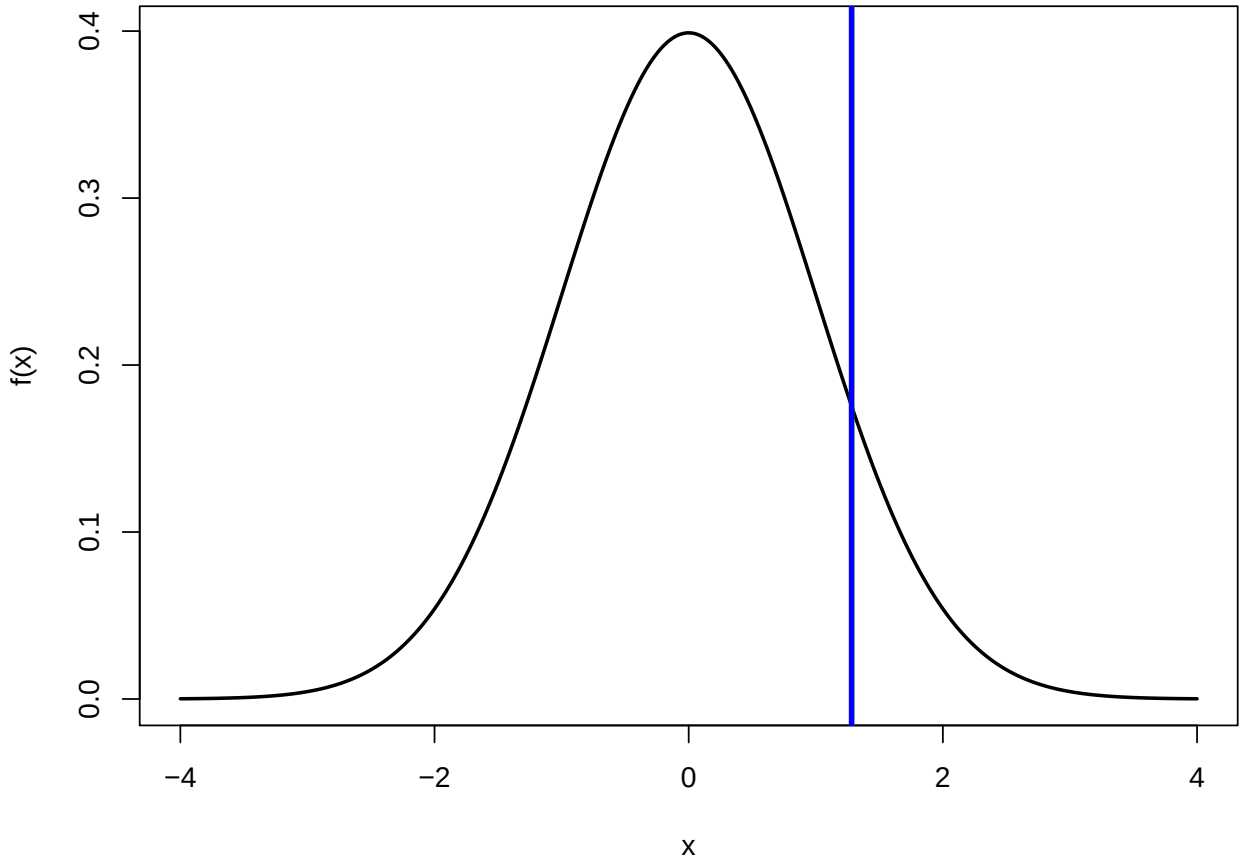
- Upper cutoff: $\Phi^{-1}(1 - .1(.75)) = 1.44$
- Lower Cutoff: $\Phi^{-1}(.1(.25)) = -1.96$
- **Tail area is $\eta_m = .1$!**

Example



- Upper cutoff: $\Phi^{-1}(1 - .1(.9)) = 1.34$
- Lower Cutoff: $\Phi^{-1}(.1(.1)) = -2.32$
- **Tail area is $\eta_m = .1$!**

Example



- Upper cutoff: $\Phi^{-1}(1 - .1(1)) = 1.28$
- Lower Cutoff: $\Phi^{-1}(.1(0)) = -\infty$
- **Tail area is $\eta_m = .1$!**

Oracle Decision Function

- **For any h_m in [0,1]**

$$\delta_m(x_m;\eta_m) = \begin{cases} 1 & \text{if} \quad x_m > \Phi^{-1}(1-\eta_m[1-h_m]) \\ 1 & \text{if} \quad x_m < \Phi^{-1}(\eta_m h_m) \\ 0 & \quad \textit{otherwise} \end{cases}$$

- Optimal: $h_m(\mu_m) = I(\mu_m < 0)$
- **BUT WE DON'T KNOW μ_m!!!**

Implementation

$$\begin{bmatrix} X_{11} & X_{12} & \dots & X_{1N} \\ X_{21} & X_{22} & \dots & X_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ X_{M1} & X_{M2} & \dots & X_{MN} \end{bmatrix} \rightarrow \begin{bmatrix} Y_1 & Z_1 \\ Y_2 & Z_2 \\ \vdots & \vdots \\ Y_M & Z_M \end{bmatrix}$$

- $\delta_1(Y_1, Y_2, \dots, Y_M, Z_1; \eta_1)$
- Plug $h_1(Y_1, Y_2, \dots, Y_M) = I(\widehat{\mu_1 < 0})$
- Z_1 is test data (rather than X_1)

Implementation

$$\begin{bmatrix} X_{11} & X_{12} & \dots & X_{1N} \\ X_{21} & X_{22} & \dots & X_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ X_{M1} & X_{M2} & \dots & X_{MN} \end{bmatrix} \rightarrow \begin{bmatrix} Y_1 & Z_1 \\ Y_2 & Z_2 \\ \vdots & \vdots \\ Y_M & Z_M \end{bmatrix}$$

- $\delta_2(Y_1, Y_2, \dots, Y_M, Z_2; \eta_2)$
- Plug $h_2(Y_1, Y_2, \dots, Y_M) = I(\widehat{\mu_2 < 0})$
- Z_2 is test data (rather than X_2)

Implementation

$$\begin{bmatrix} X_{11} & X_{12} & \dots & X_{1N} \\ X_{21} & X_{22} & \dots & X_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ X_{M1} & X_{M2} & \dots & X_{MN} \end{bmatrix} \rightarrow \begin{bmatrix} Y_1 & Z_1 \\ Y_2 & Z_2 \\ \vdots & \vdots \\ Y_M & Z_M \end{bmatrix}$$

- $\delta_M(Y_1, Y_2, \dots, Y_M, Z_M; \eta_1)$
- Plug $h_M(Y_1, Y_2, \dots, Y_M) = I(\widehat{\mu_M < 0})$
- Z_M is test data (rather than X_M)

P-values

- The compound p-value:
  - Estimate $I(\mu_m < 0)$ via $h_m(\boldsymbol{Y})$, where $\boldsymbol{Y} = (Y_1, Y_2, ..., Y_M))$

$$P(\boldsymbol{Y}, Z_m) = \min\left\{\frac{\Phi(Z_m)}{h_m(\boldsymbol{Y})}, \frac{1-\Phi(Z_m)}{1-h_m(\boldsymbol{Y})}\right\}$$

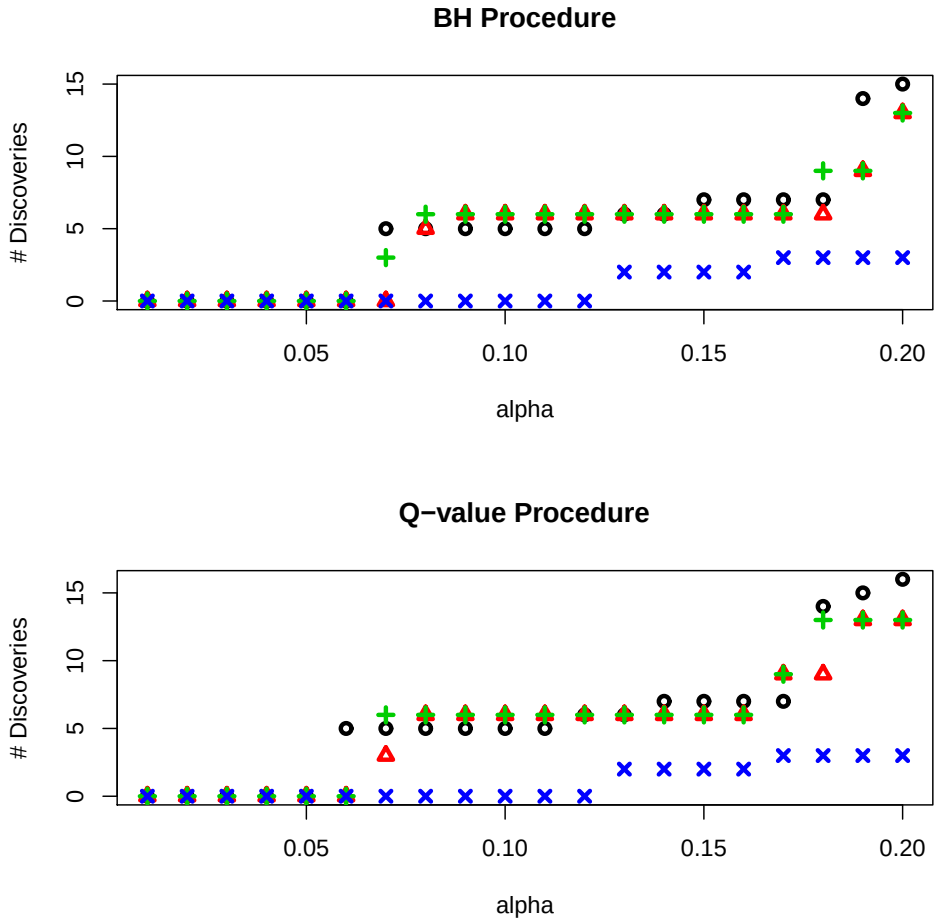
- Simple p-value:

$$P(X_m) = 2[1-\Phi(|X_m|)] = \left\{\frac{\Phi(X_m)}{1/2}, \frac{1-\Phi(X_m)}{1-1/2}\right\}$$

- The **simple** p-value uses all the data ($X_m = Y_m + Z_m$) as test data but uses $h_m = 1/2$

Standard vs. Compound *p*-values

Applied *Q*-value and BH FDR-controlling procedures to different *p*-values.



- Simple: **x** - take $h_m = .5$, use all data as test data
- Compound:**+**, **o**, △ - use 4 microarrays to estimate h_m

Analytical Results

Theorem (simplified): If test data are independent and "correctly modeled" under nulls, these compound p-values are independent and uniformly distributed under nulls.

- "Correctly Modeled" - use T-test and $\Phi^{-1}(F_T(T_m))$ if data are normal to get Z_ms, use nonparametric test otherwise. See Habiger and Pena (2011), JNS.
- Implies many multiple testing procedures will be valid

Information Borrowing Efficiency

We have ***borrowed information across tests*** to refine our upper/lower tailed cutoffs (via $h_m(\boldsymbol{Y})$)

- Gain in power related to "similarity" of data under H_{1m}s
  - *Theorem: If* $\mu_1 = \mu_2 = ...$ *under* H_{1m}*s,* $h_m(\boldsymbol{Y}) \to I(\mu_m < 0)$ *as* $M \to \infty$
  - As μ_m's become dispersed, less power is gained.

Compared to Other Approaches

Other Approaches: Efron (2001), JASA; Sun and Cai (2007), JASA, etc.

1. **We allow each δ_m to use different cutoffs depending on $f_1(z|\mu_m)$ under H_{m1}**
   - vs. using same cutoffs based on $f_1(z) = \int f_1(z|\mu)\pi(\mu)d\mu$
   - Is this better?
2. **We split sample. Others double dip.**
   - efficiency vs. validity
3. **We utilize the *p*-value**