# Some Issues of Optimality in Multiple Hypotheses Testing

**Edsel A. Peña**

(E-Mail: pena@stat.sc.edu)

Department of Statistics, University of South Carolina

Columbia, SC 29208

Some portions joint with **Josh Habiger** and **Alex McLain**

Research Support from US NIH Grants

Talk@Duke.Oct08

# Multiple Hypotheses Testing

- Multiple hypotheses testing problems (MHTP); *large* $M$*, small* $n$ settings; microarrays, proteomics, astronomy, other areas.
- Recent papers in *Annals of Stat; Stat Science; JASA*, notably Efron's. Books, e.g., Dudoit and van der Laan (2007). Well-known researchers involved on this!
- Many MHTP procedures start with $p$-values of the tests for the $M$ pairs of null and alternative hypotheses. *Example:* Benjamini-Hochberg (JRSS B, '95) FDR-controlling procedure.
- Active and challenging area with many challenges: conceptual (frequentist vs Bayes), theoretical (distributions), and computational (particle filters?).

# Mathematical Setting

Table 1: Tabular Form of Elements in an MHTP.

| 'Genes' | 1 | 2 | ... | $M$ |
|---|---|---|---|---|
| Observable Vectors (Data) | $X_1$ | $X_2$ | ... | $X_M$ |
| Data Spaces | $\mathcal{X}_1$ | $\mathcal{X}_2$ | ... | $\mathcal{X}_M$ |
| Null Hypotheses | $H_{10}$ | $H_{20}$ | ... | $H_{M0}$ |
| Alternative Hypotheses | $H_{11}$ | $H_{21}$ | ... | $H_{M1}$ |
| True States (Unknown) | $\theta_1$ | $\theta_2$ | ... | $\theta_M$ |
| Test Functions | $\delta_1$ | $\delta_2$ | ... | $\delta_M$ |
| $P$-Values | $P_1$ | $P_2$ | ... | $P_M$ |

**Note:** Each $X_m$ could be of a complicated structure, and they need not be of the same structure.

# Usual Assumptions

- $\theta_m = I\{H_{m1} \text{ is true}\}$: indicates whether $H_{m1}$ is true.
- $P_m|H_{m0} \sim U[0,1]$ and $P_m|H_{m1} \overset{st}{\leq} U[0,1]$.
- $\delta_m(x_m) \in \{0,1\}$, i.e., nonrandomized. The test $\delta_m : \mathcal{X}_m \rightarrow \{0,1\}$ depends only on $X_m$.
- Usually $\delta_m$ is chosen to be the 'best' test (MP, UMP, UMPU) when dealing with $H_{m0}$ versus $H_{m1}$ only, for each $m$.
- Generally, the $X_m$s are tacitly assumed continuous and the tests (or the $X_m$s) are independent.
- Continuity needed for uniformity of $P$-values to hold under the null hypotheses.

# Example: Two-Groups, $M$ 'Genes'

| 'Genes' | Group 1 (Control) | Group 2 (Diseased) |
|---|---|---|
| 1 | $X_{11}, \ldots, X_{1n_{11}}$ | $Y_{11}, \ldots, Y_{1n_{12}}$ |
| 2 | $X_{21}, \ldots, X_{2n_{21}}$ | $Y_{21}, \ldots, Y_{2n_{22}}$ |
| $\vdots$ | $\vdots$ | $\vdots$ |
| M | $X_{M1}, \ldots, X_{Mn_{M1}}$ | $Y_{M1}, \ldots, Y_{Mn_{M2}}$ |

May have, for example:

$$X_{m1}, X_{m2}, \ldots, X_{mn_{m1}} \text{ IID } F_m$$

$$Y_{m1}, Y_{m2}, \ldots, Y_{mn_{m2}} \text{ IID } G_m$$

$$X \perp Y$$

For each $m$: $H_{m0} : F_m = G_m$ vs $H_{m1} : F_m \overset{st}{<} G_m$

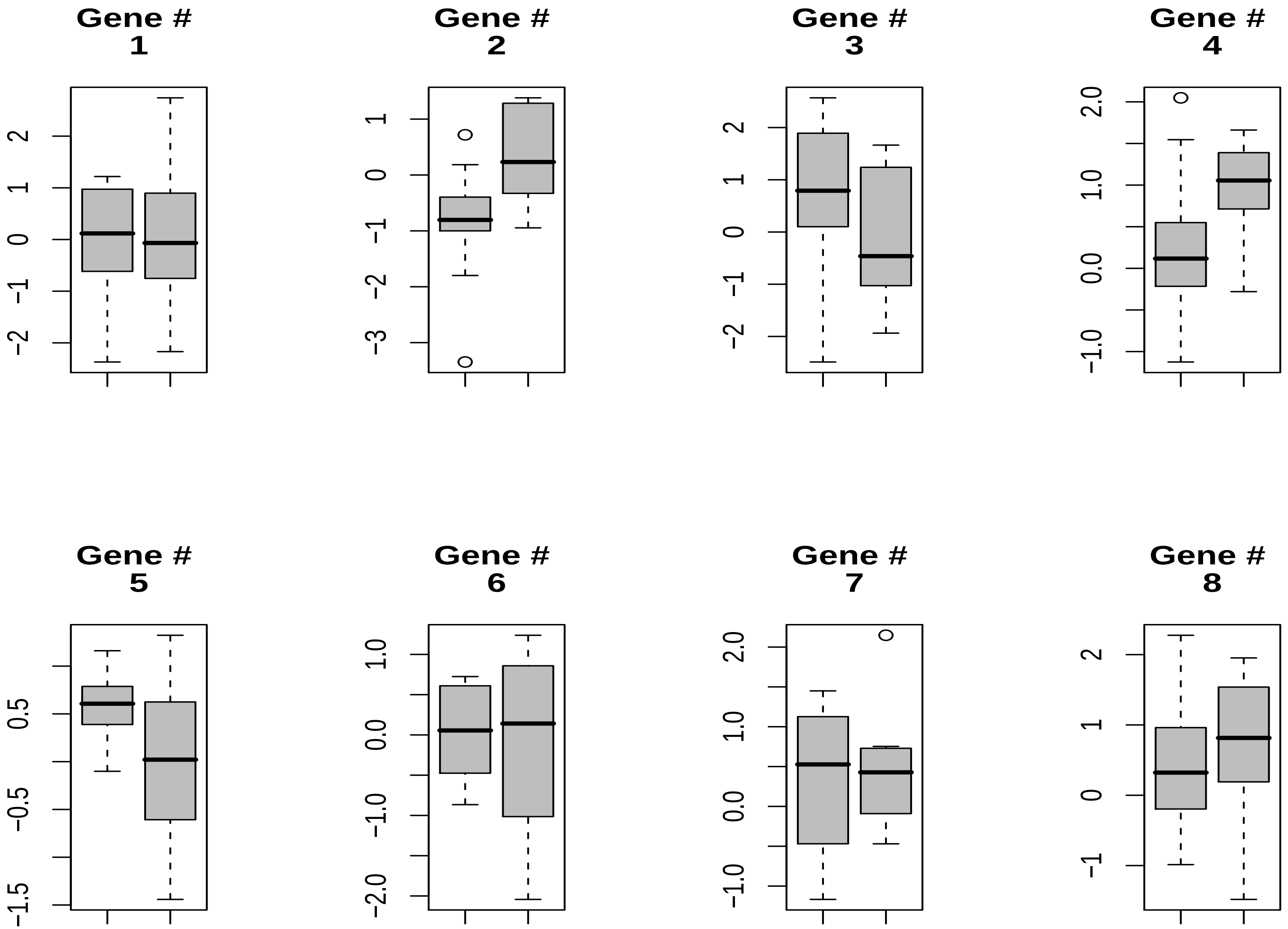
# Components of Data Set

- Number of replications, the $n_{mg}, g = 1, 2,$ need not be identical; usually $n_{mg} << M$.
- Distributions of $X$ and $Y$ may vary from gene to gene.
- For some $m$, $X$ and $Y$ may be discrete, in others continuous.
- For each gene data could be a multi-group data, a regression-type data, or of more complicated form.
- Problem: Determine genes for which the distributions of $X$ and $Y$ differ. Thus, to test $M$ pairs of null and alternative hypotheses, $H_{m0}$ vs $H_{m1}$.
- Issues: Presentation and visualization of data. Also, efficient reduction to lower-dimensional spaces.

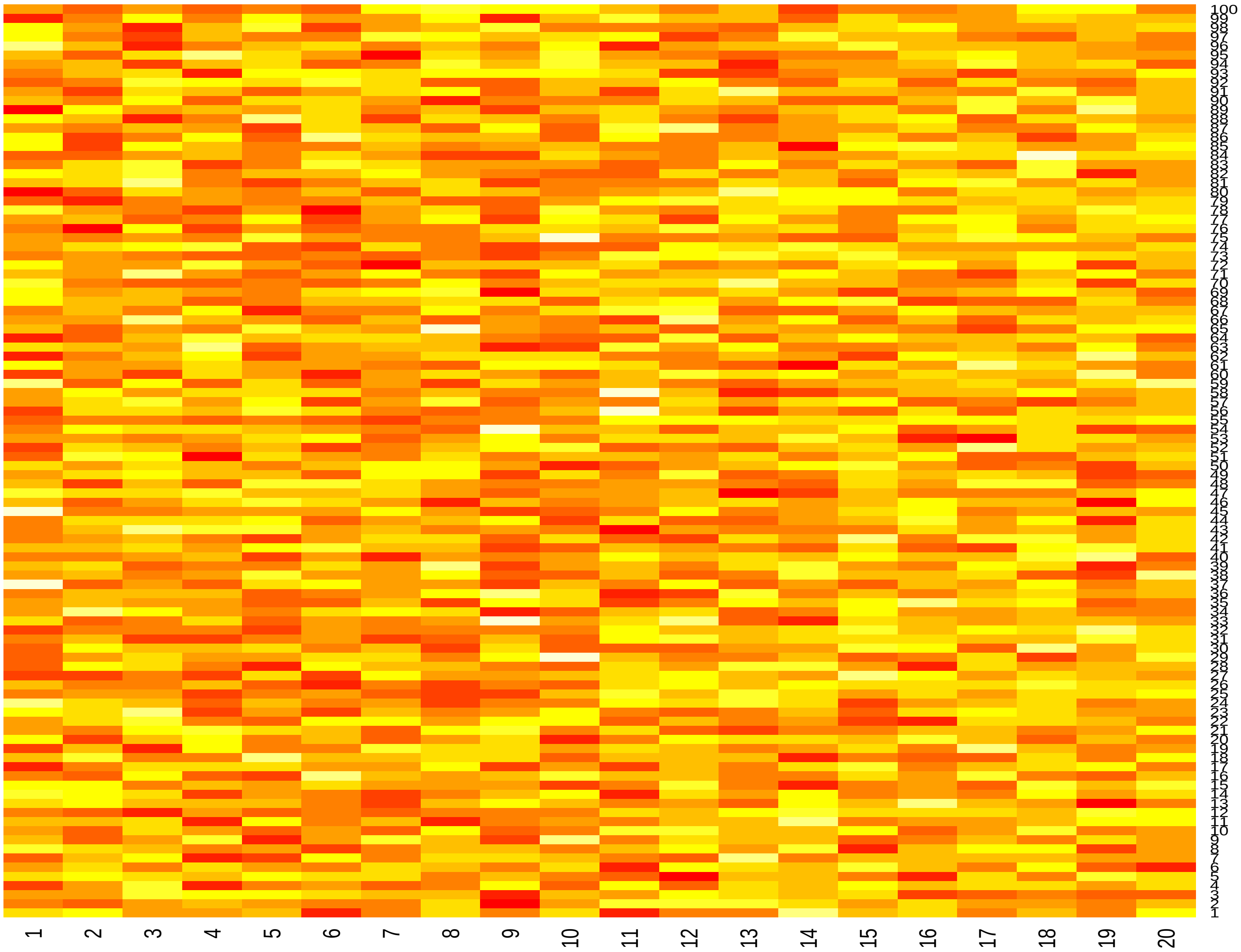
# Illustration: Simulated Data

- Data Generation
- $G = 2$ groups; $M = 100$ 'genes'
- $\theta_1, \theta_2, \ldots, \theta_m$ IID $Ber(0.20)$. These are the indicators of which alternative hypotheses are correct. 80% chance of a correct null.
- $\eta_m = |$Normal$(2,1)|I\{\theta_m = 1\}$. These are the true alternative means. If non-zero, alternative is correct.
- $n_1 = n_2 = 10$: number of replications per group per gene.
- $X_{mj}, j = 1, 2, \ldots, n$ are IID Normal$(0,1)$
- $Y_{mj}, j = 1, 2, \ldots, n$ are IID Normal$(\eta_m, 1)$

# Visualizations: Paired Box Plots, First 8 Genes

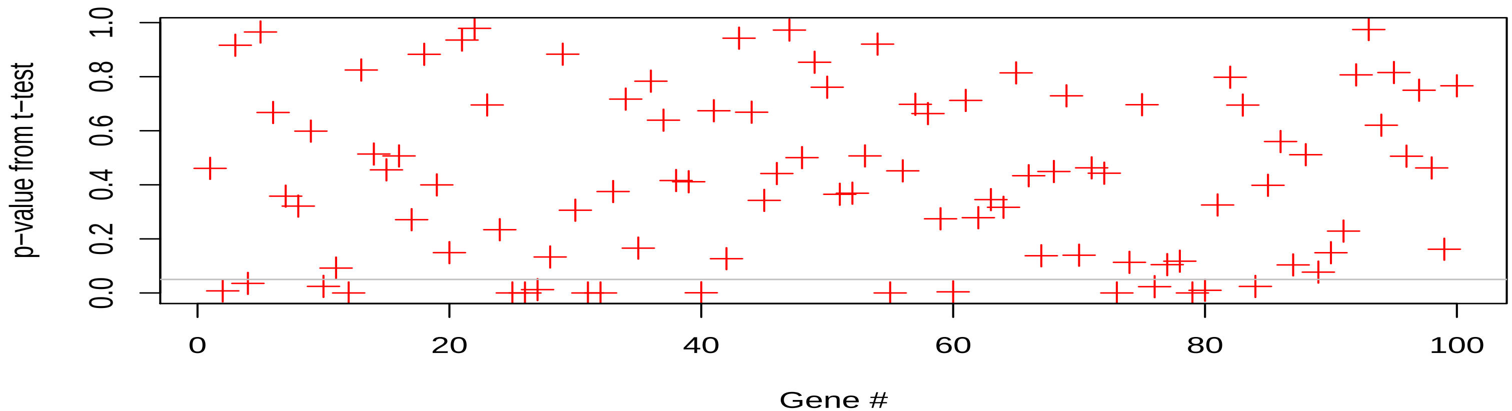


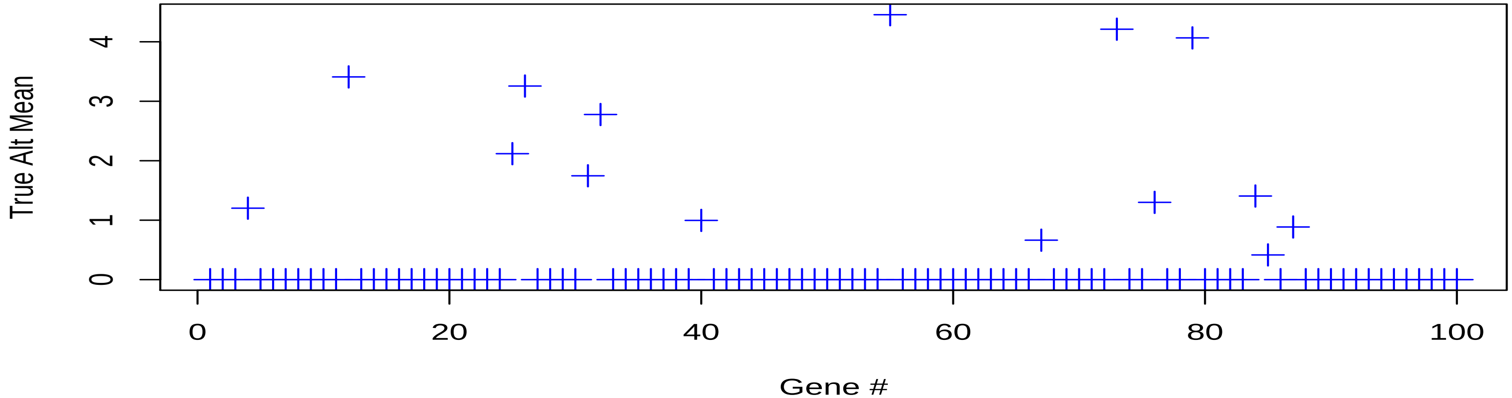
# Visualizations: A Heat Map



# Decision-Making

$P$-Values are from the two-sample $t$-test for each gene.



# ‘Unconscious’ Statistician’s Rule

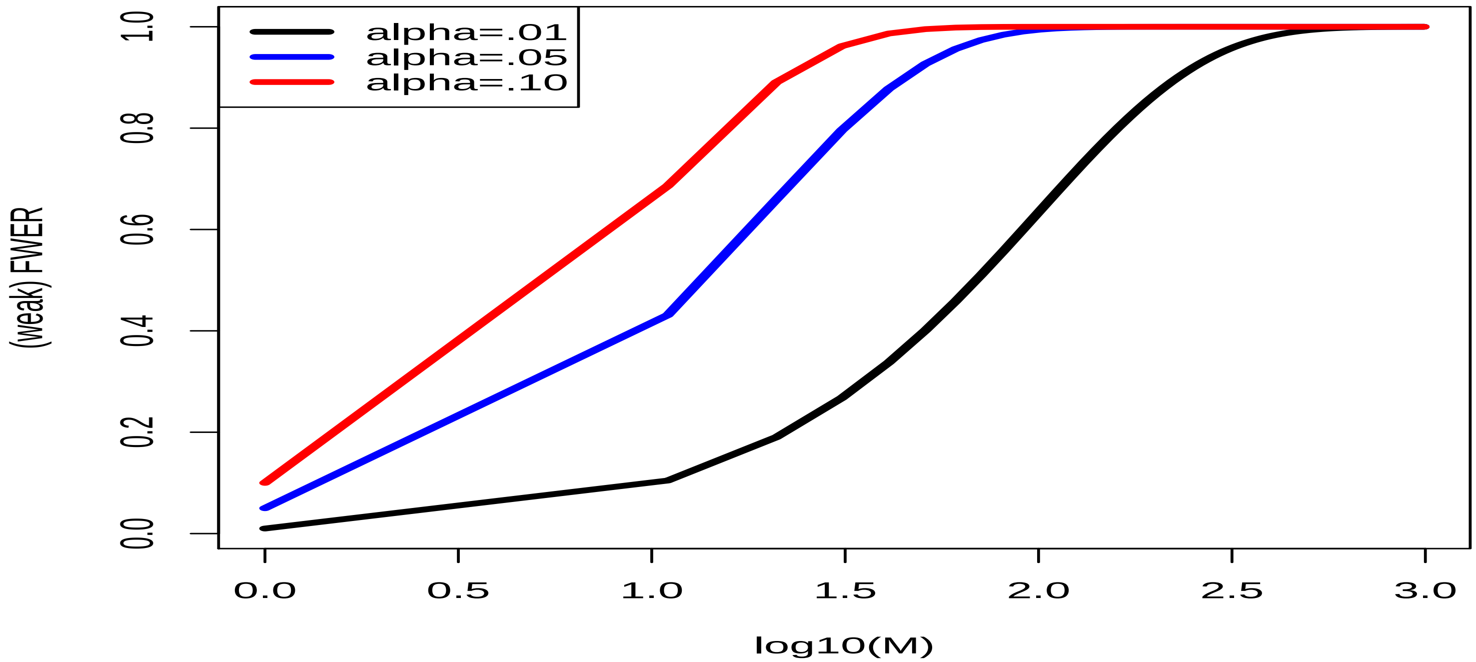
*Decision Rule:* Reject all $H_{m0}$s with $P_m \leq 0.05$.

- A summary of the performance:

| | Hypotheses Accepted | Hypotheses Rejected | Total |
|---|---|---|---|
| Correct Nulls | $U = 80$ | $V = 5$ | $M_0 = 85$ |
| False Nulls | $T = 3$ | $S = 12$ | $M_1 = 15$ |
| Total | $M - R = 83$ | $R = 17$ | $M = 100$ |

- Family-wise error rate (FWER) = 100%.
- False Discovery Rate (FDR) = (5/17)*100 = 29%.
- Missed Discovery Rate (MDR) = (3/15)*100 = 20%.

# ‘Jekyll & Hyde’ of Multiplicity




- *Random Typewriter* duplicates Dan Brown’s *Angels and Demons*, not only once, but twice!
- Someone wins a multimillion lottery twice!

# Formalization: Spaces and Losses

- Parameter ($\theta$) Space: $\Theta = \{0,1\}^M$
- Action ($a$) Space: $\mathcal{A} = \{0,1\}^M$
- Data ($x$) Space: $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_M$
- $L_0(a,\theta) = I\left\{\sum_{m=1}^M a_m(1-\theta_m) > 0\right\}$
- $L_1(a,\theta) = \left[\frac{\sum_{m=1}^M a_m(1-\theta_m)}{\sum_{m=1}^M a_m}\right] I\left\{\sum_{m=1}^M a_m > 0\right\}$
- $L_2(a,\theta) = \left[\frac{\sum_{m=1}^M (1-a_m)\theta_m}{\sum_{m=1}^M \theta_m}\right] I\left\{\sum_{m=1}^M \theta_m > 0\right\}$
- Note that $L_1(a,\theta)$ is the false discovery rate (FDR) and $L_2(a,\theta)$ is the missed discovery rate (MDR) for action $a$ and state $\theta$.

# Decision and Risk Functions

- MHTP Decision Function (MHTPDF):

$$\delta = (\delta_1, \delta_2, \ldots, \delta_M) : \mathcal{X} \rightarrow \mathcal{A}$$

Risk Functions for a MHTPDF $\delta$

- $R_0(\delta, \theta) = E_\theta[L_0(\delta(X), \theta)]$.
- FWER$(\delta) \equiv R_0(\delta, \mathbf{0})$, family-wise error rate.
- $R_1(\delta, \theta) = E_\theta[L_1(\delta(X), \theta)]$, (expected) FDR.
- $R_2(\delta, \theta) = E_\theta[L_2(\delta(X), \theta)]$, (expected) MDR.

# FWER Control: Sidak Procedure

- Given $\alpha \in (0, 1)$, define

$$\eta = 1 - (1 - \alpha)^{1/M}.$$

- The Sidak MHTPDF rejects all null hypothesis $H_{m0}$ with $p_m(x_m) \leq \eta$, where $p_m(x_m)$ is the observed $p$-value for testing $H_{m0}$ versus $H_{m1}$.
- Procedure is $p$-value based.
- Independence of the $X_m, m = 1, 2, \ldots, M$, crucially needed to achieve control.

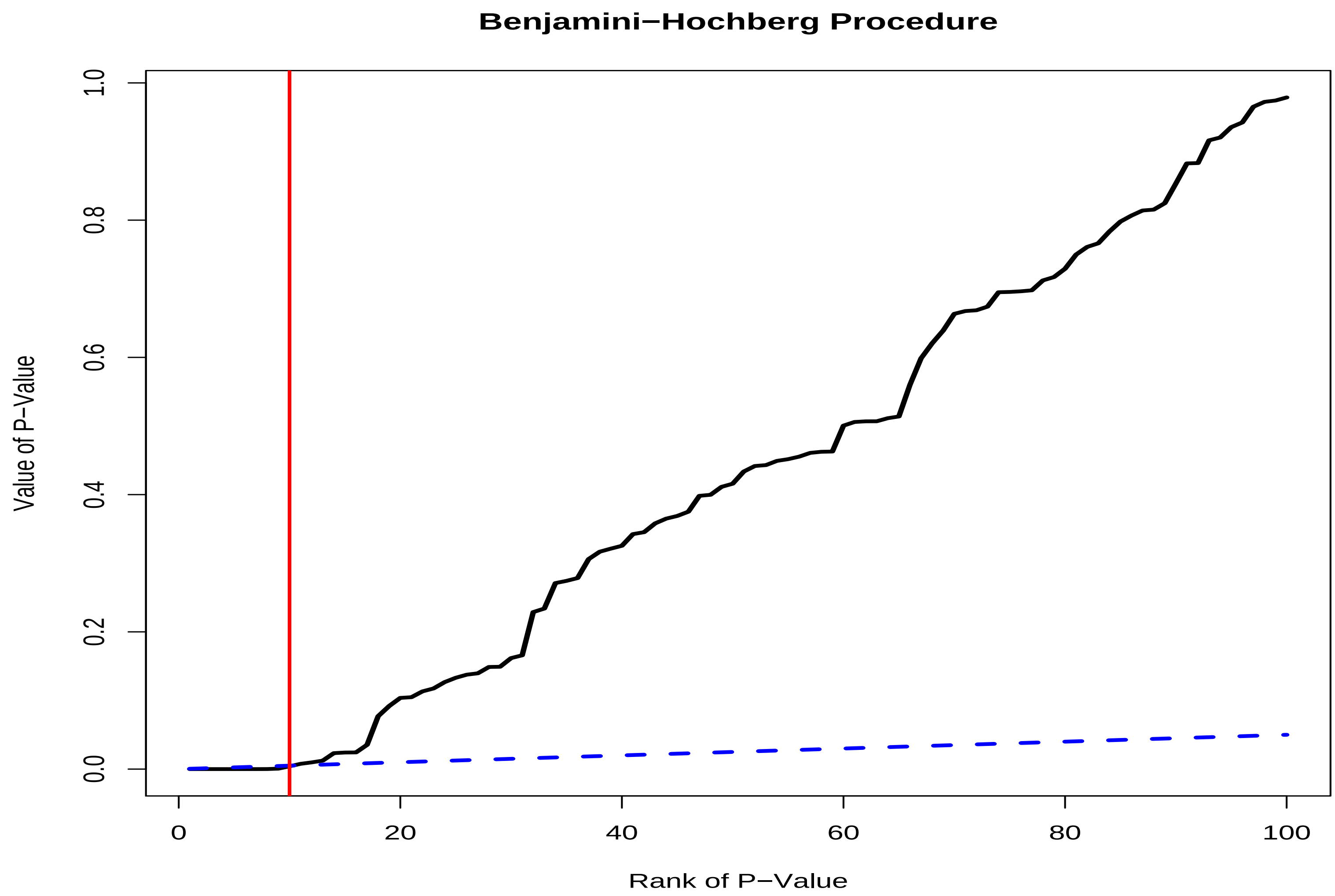
# FDR Control: BH Procedure

- Let $q^* \in (0,1)$ be the desired FDR level.
- Let $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(M)}$ be the ordered $p$-values, and let $H_{(m)0}$ be the null hypothesis associated with $p_{(m)}$. Define

$$J = \max\left\{m \in \{1, 2, \ldots, M\} : \ p_{(m)} \leq \frac{q^* m}{M}\right\}.$$

- BH MHTPDF: Reject all $H_{(m)0}$ for $m = 1, 2, \ldots, J$.
- Benjamini-Hochberg (JRSS B (95)) proved that this $p$-value based procedure, which is adaptive, achieves the desired FDR control at $q^*$ whatever $\theta_0$ is.

# BH Plot on Illustrative Data



# Sidak & BH Performances: $\alpha = .05$

| | Hypotheses Accepted | Hypotheses Rejected | Total |
|---|---|---|---|
| Correct Nulls | 85, 84 | 0, 1 | 85 |
| False Nulls | 7, 6 | 8, 9 | 15 |
| Total | 92, 90 | 8, 10 | 100 |

- Observed Error Rates:
- Sidak: FWER = 0; FDR = 0; MDR = (7/15)*100 = 47%.
- BH: FWER = 100%; FDR = (1/10)*100 = 10%; MDR = (6/15)*100 = 40%.

# The Motivating Question

- What is the *role* of the power functions of the individual tests in MHTP procedures, or did we use them at all?
- FWER-controlling procedures, e.g., Sidak procedure, or FDR-controlling procedures, e.g., BH procedure, assumes the same powers for each of the $M$ tests as the $p$-values are treated in a *symmetric fashion*.
- Unlikely however that $M$ tests will all have the same powers.
- Different power functions may arise due to different distributions, tests used (t-test; Wilcoxon), or effect sizes.

# A Look into History

- 1920-30s, Neyman and Pearson: consider alternatives. Contrast to then-existing significance testing ($p$-value) approach.
- NP framework: most powerful (MP) and uniformly most powerful (UMP) tests; monotone likelihood ratio (MLR) property.
- In MHTP, view configurations of $M$ pairs of hypotheses as 'alternative.' From NP lesson, behooves to exploit alternative configuration and individual powers of the tests.
- As in NP theory, start with simple versus simple hypotheses per gene!

# Revised Mathematical Setting

| | | | | |
|---|---|---|---|---|
| 'Genes' | 1 | 2 | ... | $M$ |
| Observed Data | $X_1$ | $X_2$ | ... | $X_M$ |
| Data Spaces | $\mathcal{X}_1$ | $\mathcal{X}_2$ | ... | $\mathcal{X}_M$ |
| Density of $X_m$ | $f_1$ | $f_2$ | ... | $f_M$ |
| *Randomizers* | $U_1$ | $U_2$ | ... | $U_M$ |
| Nulls | $H_{10} : f_{10}$ | $H_{20} : f_{20}$ | ... | $H_{M0} : f_{M0}$ |
| Alternatives | $H_{11} : f_{11}$ | $H_{21} : f_{21}$ | ... | $H_{M1} : f_{M1}$ |
| True States | $\theta_1$ | $\theta_2$ | ... | $\theta_M$ |
| NP MP Tests | $\delta_1^*(\eta_1)$ | $\delta_2^*(\eta_2)$ | ... | $\delta_M^*(\eta_M)$ |
| Test Sizes | $\eta_1$ | $\eta_2$ | ... | $\eta_M$ |
| Test Powers | $\pi_1(\eta_1)$ | $\pi_2(\eta_2)$ | ... | $\pi_M(\eta_M)$ |

# Elements of Revised Setting

- $f_{m0}$: known density or mass functions.
- $f_{m1}$: known density or mass functions.
- $U_1, U_2, \ldots, U_M$ are IID $U[0,1]$ variables, independent of the $X_m$s.
- $U_m$s auxiliary data generated at start of experiment. Used only if there is a need to randomize in each of the tests.
- $\delta_m^*(X_m, U_m; \eta_m)$ is the *non*randomized (we have a randomizer $U_m$) Neyman-Pearson most powerful test for $H_{m0}$ vs $H_{m1}$ of size $\eta_m$.
- $\pi_m(\eta_m) = \Pr\{\delta_m(X_m, U_m; \eta_m) = 1 | X_m \sim f_{m1}\}$: power of test $\delta_m(\eta_m)$. Viewed as a function of the size $\eta_m$.

# Students, Quick! NP MP-Test

For testing $H_{m0} : f_m = f_{m0}$ versus $H_{m1} : f_m = f_{m1}$ based on $X_m$, the size $\eta_m$ most powerful test is of form:

$$\delta_m(X_m;\eta_m) = \begin{cases} 1 & \text{if} \quad \lambda_m(X_m) > c_m(\eta_m) \\ \gamma_m(\eta_m) & \text{if} \quad \lambda_m(X_m) = c_m(\eta_m) \\ 0 & \text{if} \quad \lambda_m(X_m) < c_m(\eta_m) \end{cases},$$

where

$$\lambda_m(x_m) = \frac{f_{m1}(x_m)}{f_{m0}(x_m)}$$

and $c_m(\eta_m)$ and $\gamma_m(\eta_m) \in [0,1)$ are chosen to satisfy the size requirement $E\{\delta_m(X_m;\eta_m)|X_m \sim f_{m0}\} = \eta_m$.

# Using the Randomizer $U_m$

The NP most powerful test may need to randomize when $\lambda_m(x_m) = c_m(\eta_m)$. As we statisticians are apt to proclaim,

When in doubt, Randomize!

When given the auxiliary data $U_m$, it could be made a nonrandomized test via:

$$\delta_m^*(X_m, U_m; \eta_m) = I\{\delta_m(X_m; \eta_m) = 1\} + I\{\delta_m(X_m; \eta_m) = \gamma_m(\eta_m);\ U_m \leq \gamma_m(\eta_m)\}.$$

This is the form of the tests displayed in the table of the revised mathematical setting.

# (Optimal) Choice of MHTPDF $\delta$

- **With FWER-Control at Level $\alpha$:**

  Given an $\alpha \in (0,1)$, to find a $\delta$ such that FWER$(\delta) = R_0(\delta, \mathbf{0}) \leq \alpha$ with $R_2(\delta, \mathbf{1})$ minimized (or made small).

- **With FDR-Control at Level $q^*$:**

  Given a $q^* \in (0,1)$, to find a $\delta$ such that $R_1(\delta, \theta_0) \leq q^*$ with $R_2(\delta, \mathbf{1})$ minimized (or made small). Here, $\theta_0$ is the true state and is *unknown*.

# (Weak) FWER and MDR

Suppose then that the respective sizes of the MP tests are $\eta_1, \eta_2, \ldots, \eta_M$. Then,

$$\text{FWER}(\delta^*) = 1 - \prod_{m=1}^{M} (1 - \eta_m);$$

and

$$R_2(\delta^*, \mathbf{1}) = \frac{1}{M} \sum_{m=1}^{M} (1 - \pi_m(\eta_m)).$$

# Optimal FWER Control

The problem of choosing an MHTPDF with FWER $\leq \alpha$ amounts therefore to choosing the test sizes

$$(\eta_1(\alpha), \eta_2(\alpha), \ldots, \eta_M(\alpha))$$

such that

$$\sum_{m=1}^{M} \pi_m(\eta_m) \text{ is maximized}$$

subject to the constraint

$$\prod_{m=1}^{M} (1 - \eta_m) \geq 1 - \alpha.$$

# Existence

*Theorem:* For any $\alpha \in (0,1)$, there always exists a size vector

$$\eta(\alpha) = (\eta_1(\alpha), \eta_2(\alpha), \ldots, \eta_M(\alpha))$$

that solves the constrained optimization problem.

Hence an optimal MHTPDF that controls the FWER among the (restricted) class of decision functions always exists.

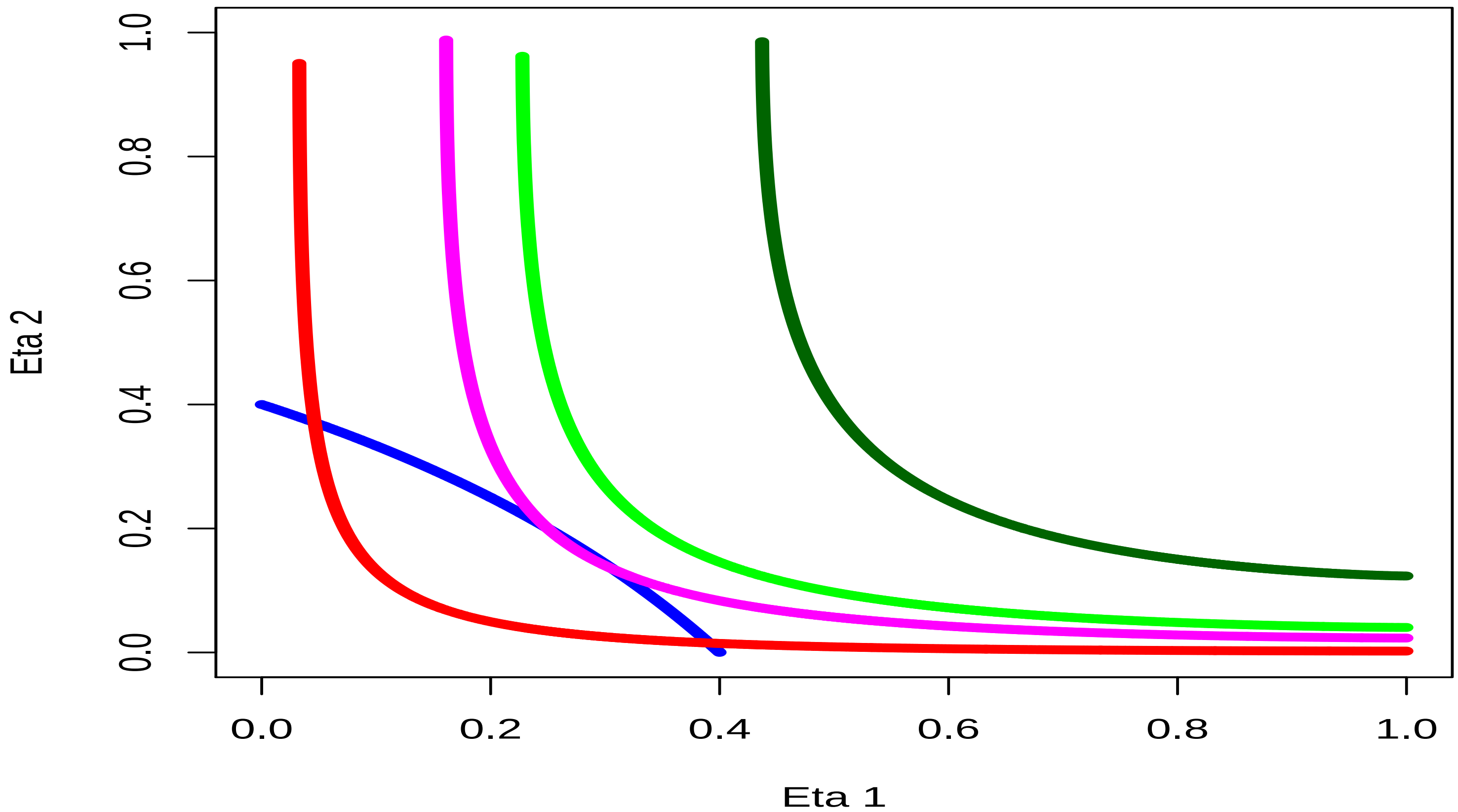
*Remark:* Restricted class since $\delta_m$ is made only to depend on $X_m$ for each $m$.

# Main Ideas Behind Proofs

- $\eta_m \mapsto \pi_m(\eta_m)$ is a concave, continuous, and nondecreasing function, with $\pi_m(1) = 1$.
- The constraint set $C_\alpha = \{\eta : \prod_m (1 - \eta_m) \geq 1 - \alpha\}$ is a closed and convex set containing $\mathbf{0}$.
- For each $b$, the set $\mathcal{N}_b = \{\eta : \sum_m \pi_m(\eta_m) \geq Mb\}$ is a closed and convex set containing $\mathbf{1}$ and is nonincreasing in $b$. Also, $\mathcal{N}_0 = [0, 1]^M$.
- Maximize $b$ such that $C_\alpha \cap \mathcal{N}_b \neq \emptyset$.
- Separating Hyperplane Theorem guarantees the existence of such an optimal $b^* = b$.
- A size vector in the non-empty intersection $C_\alpha \cap \mathcal{N}_{b^*}$ is optimal.

# Case of $M = 2$: Regions in $\eta$-Space

BLUE: *Upper* Boundary of $C_\alpha$ for $\alpha = .40$; **Other Colors:** *Lower* Boundaries of $\mathcal{N}_b$ for Increasing $b$.



# Uniqueness

*Theorem:* If the power functions $\eta_m \mapsto \pi_m(\eta_m)$ are strictly increasing for each $m = 1, 2, \ldots, M$, then the optimal size vector $(\eta_1(\alpha), \eta_2(\alpha), \ldots, \eta_M(\alpha))$ is unique.

*Remark:* Cases where non-uniqueness occur are associated with non-regular families such as the uniform distribution or shifted exponential where the power function, as a function of the size, could equal one for sizes less than one.

*Corollary:* The Sidak MHTFDF obtains when the power functions $\eta_m \mapsto \pi_m(\eta_m)$ for $m = 1, 2, \ldots, M$ are identical.

# When Twice-Differentiable

- *Theorem:* If $\eta_m \mapsto \pi_m(\eta_m)$ is twice-differentiable with first derivative $\pi'_m(\eta_m)$ and second derivative $\pi''_m(\eta_m)$, the optimal size vector $(\eta_1, \eta_2, \ldots, \eta_M)$ solves the Lagrange equations

$$\forall m : \quad \pi'_m(\eta_m)(1-\eta_m) = \lambda \in \Re;$$

$$\sum_{m=1}^{M} \log(1-\eta_m) = \log(1-\alpha).$$

- In PHM (08) we have written an *R* code to compute this optimal size vector for certain situations involving normal, exponential, and binomial distributions.

# Families with MLR Property

- Formulation is for simple null vs simple alternative for each $m$ so appears limited.
- Suppose $X_m \sim f_m \in \mathcal{F}_m = \{f_m(x;\beta_m) : \beta_m \in \Re\}$ possessing monotone likelihood ratio (MLR) property.
- UMP exists for $H_{m0} : \beta_m \leq \beta_{m0}$ vs $H_{m1} : \beta_m > \beta_{m0}$.
- Focus might be on $\beta_{m1} (> \beta_{m0})$ on which a desired power is needed, and this determines *effect size*. Power is evaluated at the value $\beta_{m1}$.
- Therefore, framework extends more generally in MLR families.
- In the examples, the elements of effect size vector is varied to induce different powers.

# Example: Normal Distributions

- Setting: $X_m \sim N(\mu_m, 1), m = 1, 2, \ldots, M$.
- At each $m$, to test $H_{m0} : \mu_m \le 0$ vs $H_{m1} : \mu_m > 0$.
- The UMP test of level $\eta_m$:

$$\delta_m^*(X_m; \eta_m) = I\{X_m > \Phi^{-1}(1 - \eta_m)\}$$

  with $\Phi^{-1}(\cdot)$ is standard normal quantile function.
- Effect Size: $\gamma_m = \mu_{m1}$. Power at this effect size is

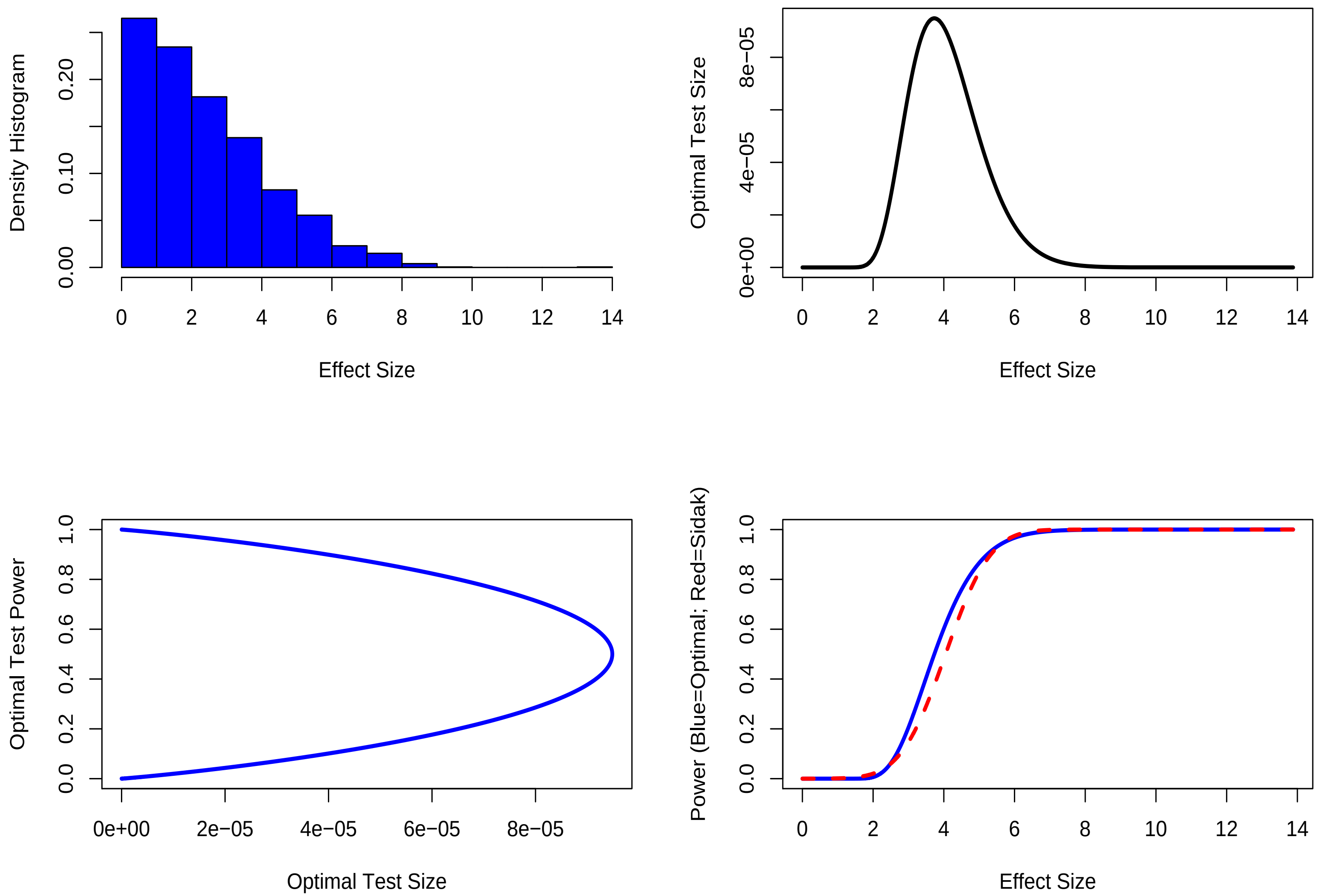
$$\pi_m(\eta_m) = 1 - \Phi(\Phi^{-1}(1 - \eta_m) - \gamma_m).$$

- Effect Size Vector: $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_M)$.

# Normal Example: Small $M$

| Effect Size, $\gamma$, Configuration | Size Vector/[Effi over Sidak] $M = 20$ |
|---|---|
| $M/2 : (.5, 1)$ | $10 : (0, .0051)$ [125.1] |
| $M/2 : (1, 5)$ | $10 : (.0035, .0016)$ [100.3] |
| $M/4 : (0.5, 1, 2, 4)$ | $5 : (0, .0003, .0068, .0031)$ [107.1] |

# Normal Example: $M = 2000$; $\gamma_m \overset{IID}{\sim} |N(0,3)|$



# Testing with Exponential DFs

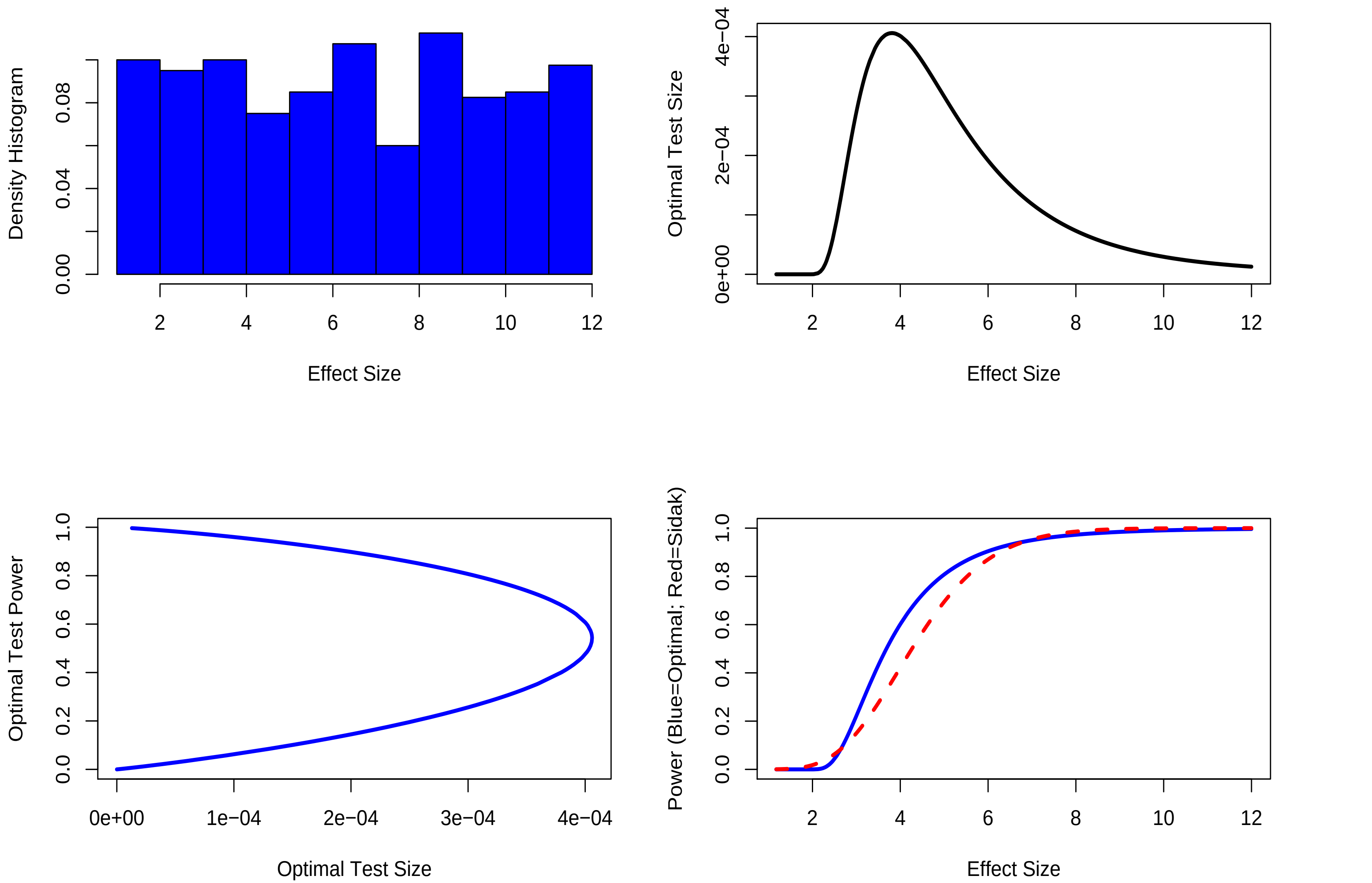
- $X_{mi}, i = 1, \ldots, n,$ IID Exp$(\lambda_m)$.
- Test: $H_{m0} : \lambda_m = \lambda_{m0}$ vs $H_{m1} : \lambda_m = \lambda_{m1} (> \lambda_{m0})$.
- Sufficient Statistics: $S_m = \sum_{i=1}^n X_{mi}$.
- NP Test of Size $\eta_m$:

$$\delta_m^{NP}(S_m; \eta_m) = I\{2\lambda_{m0} S_m \leq c_m(\eta_m)\}$$

- $c_m(\eta_m) = G_{2n}^{-1}(\eta_m)$; $G_k(\cdot)$ is $\chi_k^2$ df.
- Effect Sizes: $\rho_m = \lambda_{m1}/\lambda_{m0}$.
- Power Functions:

$$\pi_m^{NP}(\eta_m) = G_{2n}(\rho_m G_{2n}^{-1}(\eta_m))$$

# Exponential Example: $M = 400$; $\gamma_m \overset{IID}{\sim} U[1.1, 12]$; $n = 10$



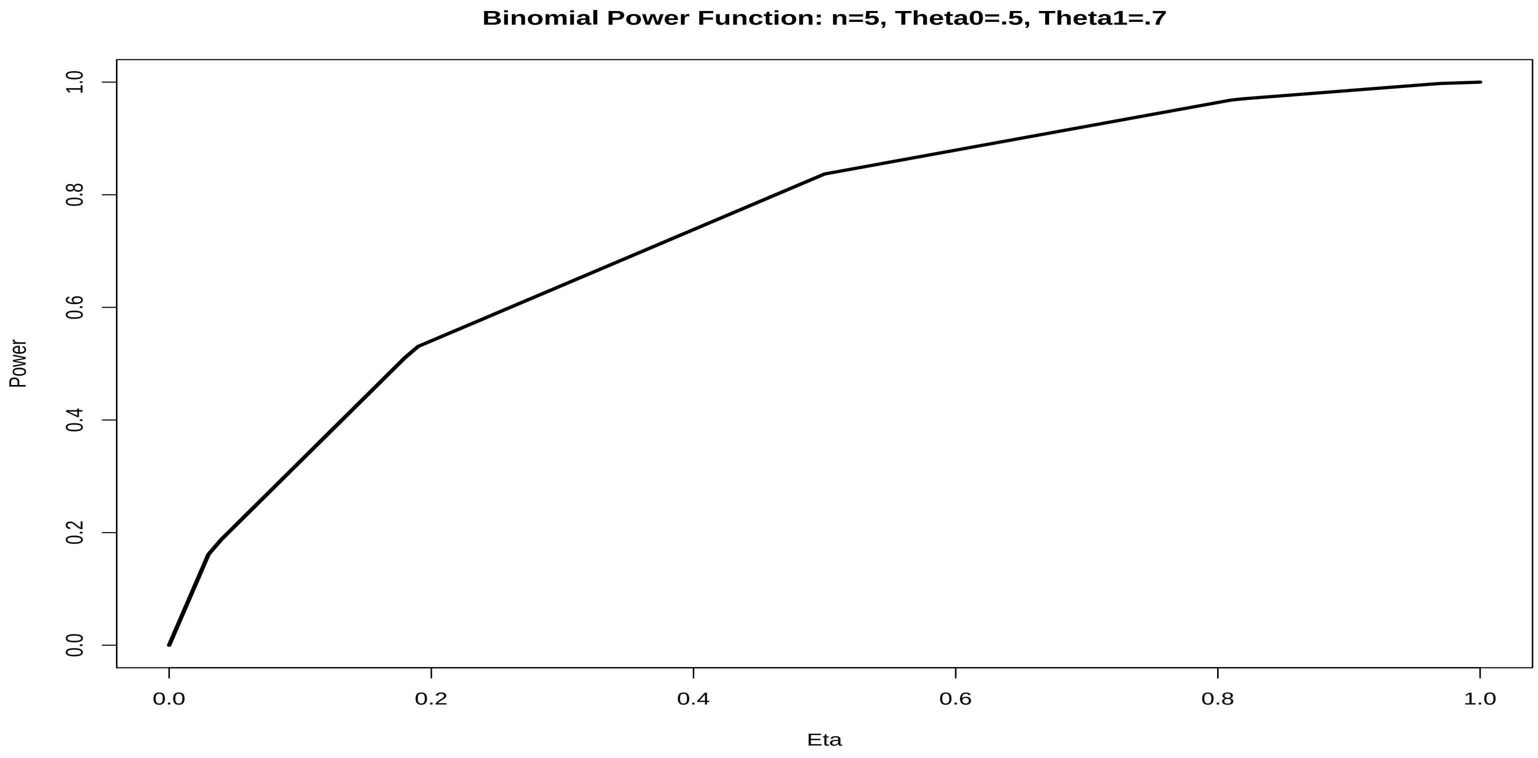
# Testing with Binomials

Set-Up: $X_m \sim B(n_m = 5, \theta_m)$; $H_{m0} : \theta_m = \theta_{m0} = .2$; $H_{m1} : \theta_m = \theta_{m1}$. $\theta_{m1}$ generated uniformly over $[.2, 1]$.

| $m$ | $\theta_{m0}$ | $\theta_{m1}$ | Optimal Size | Optimal Power | Sidak Size | Sidak Power |
|---|---|---|---|---|---|---|
| 1 | 0.2 | 0.329 | 9.55e-04 | 0.019 | 0.005 | 0.060 |
| 2 | 0.2 | 0.440 | 1.04e-02 | 0.277 | 0.005 | 0.207 |
| 3 | 0.2 | 0.599 | 9.76e-03 | 0.658 | 0.005 | 0.575 |
| 4 | 0.2 | 0.661 | 6.37e-03 | 0.775 | 0.005 | 0.722 |
| 5 | 0.2 | 0.682 | 6.36e-03 | 0.818 | 0.005 | 0.769 |
| 6 | 0.2 | 0.780 | 6.40e-03 | 0.952 | 0.005 | 0.927 |
| 7 | 0.2 | 0.795 | 6.36e-03 | 0.964 | 0.005 | 0.943 |
| 8 | 0.2 | 0.843 | 3.56e-03 | 0.965 | 0.005 | 0.977 |
| 9 | 0.2 | 0.949 | 8.64e-04 | 0.998 | 0.005 | 0.999 |
| 10 | 0.2 | 0.999 | 4.19e-06 | 0.999 | 0.005 | 1.000 |
| Total | NA | NA | NA | 7.430 | NA | 7.184 |

# Some Observations

- Computations in the binomial example more elaborate since each power function (wrt size) is polygonal (see below) hence does *not* allow the Lagrange approach.



# Continued ...

- Both the normal and exponential settings allowed the Lagrange solution approach.
- General characteristics of the optimal size vector and the powers under this optimal size vector for the binomial example are similar to the normal and exponential examples.
- Patterns similar as well when the effect sizes were generated by a non-uniform distribution.
- Observe the improvement in overall discovery rate over the Sidak procedure.

# Lesson on Investing Size

- Observe that small optimal sizes are allocated to those where effect size is either small (which converts to low power) or effect size is large (which converts to high power).
- Intuitive, in hindsight, and is indeed a size investment 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!

# Extending to FDR-Control

- The optimal FWER-controlling procedure can be extended to make it into an FDR-controlling procedure in the spirit of Benjamini-Hochberg.
- Idea is to use the FWER value $\alpha$ as the 'anchor' which will then lead to the determination of the optimal sizes for the $M$ tests.
- Let

$$\alpha \mapsto (\eta_1(\alpha), \eta_2(\alpha), \ldots, \eta_M(\alpha))$$

denote the mapping from FWER-value $\alpha$ to the $M$ tests' optimal sizes as guaranteed by the earlier results.

## Proposed Generalized BH Procedure

- Desired FDR-level: $q^*$. Define $\alpha_M^* \equiv \alpha_M^*(\mathbf{X}, \mathbf{U})$ via

$$\alpha_M^* = \sup\left\{\alpha \in (0,1) : \sum_{m=1}^M \eta_m(\alpha) \le q^* \sum_{m=1}^M \delta_m^*(X_m, U_m; \eta_m(\alpha))\right\}.$$

- The proposed FDR-controlling MHTPDF is

$$\delta^*(\alpha_M^*) = (\delta_m^*(X_m, U_m; \eta_m(\alpha_M^*)), m = 1, 2, \ldots, M).$$

- **Conjecture:** Whatever $\theta_0$ is, $R_1(\delta^*(\alpha_M^*), \theta_0) \le q^*$.

# Intuition & Motivation (Informal Proof)

$$Q_M(\delta^*(\alpha)) = \frac{\sum_m \delta_m(\eta_m(\alpha)))(1-\theta_m)}{\sum_m \delta_m(\eta_m(\alpha))}$$

$$E\left\{\sum_m \delta_m(\eta_m(\alpha)(1-\theta_m)\right\} \leq E_{\mathbf{0}}\left\{\sum_m \delta_m(\eta_m(\alpha))\right\} = \sum_m \eta_m(\alpha)$$

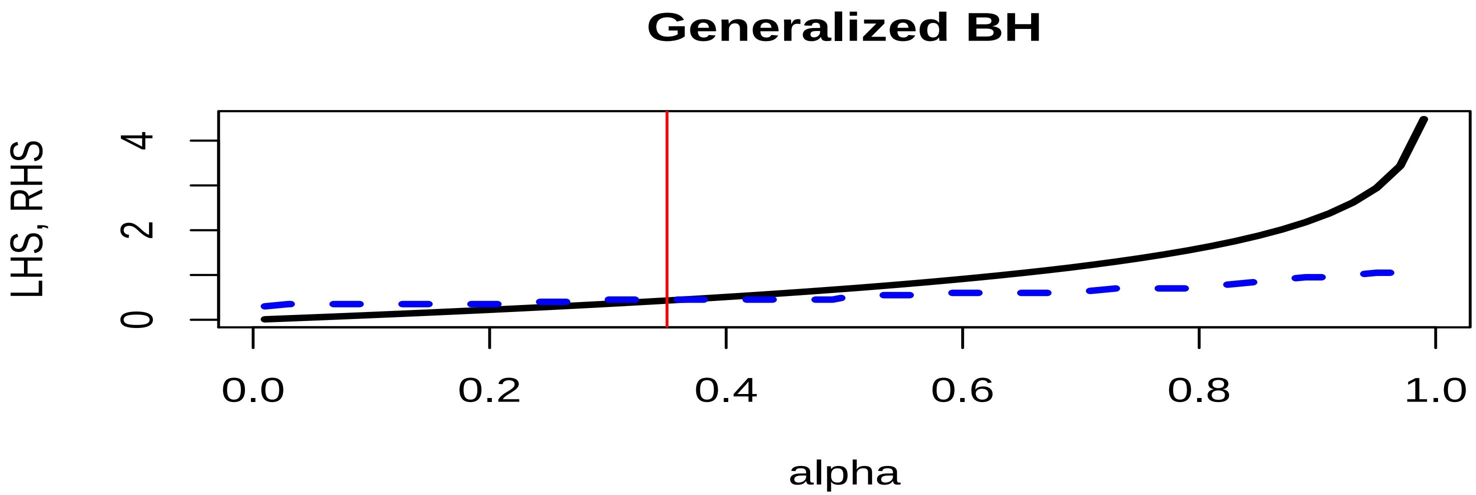
$$Q_M(\delta^*(\alpha)) \ \lesssim \ \frac{\sum_m \eta_m(\alpha)}{\sum_m \delta_m(\eta_m(\alpha))}$$

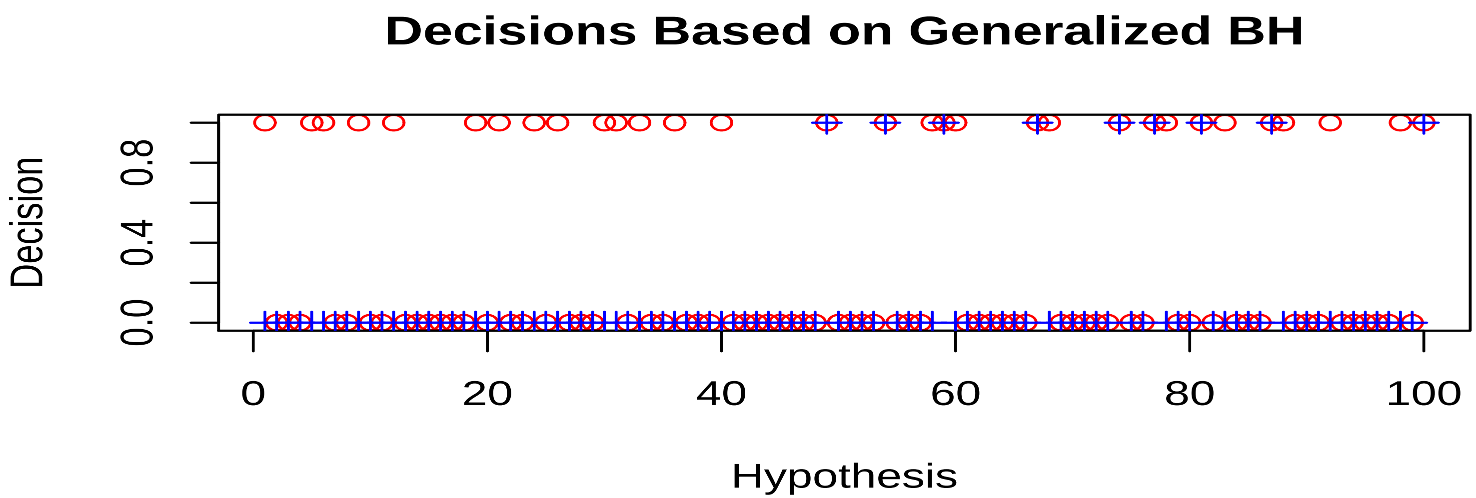
Optimize! $\alpha_M^* = \sup\left\{\alpha : \sum_m \eta_m(\alpha) \leq q^* \sum_m \delta_m(\eta_m(\alpha))\right\}$

# Illustration: On Simulated Data

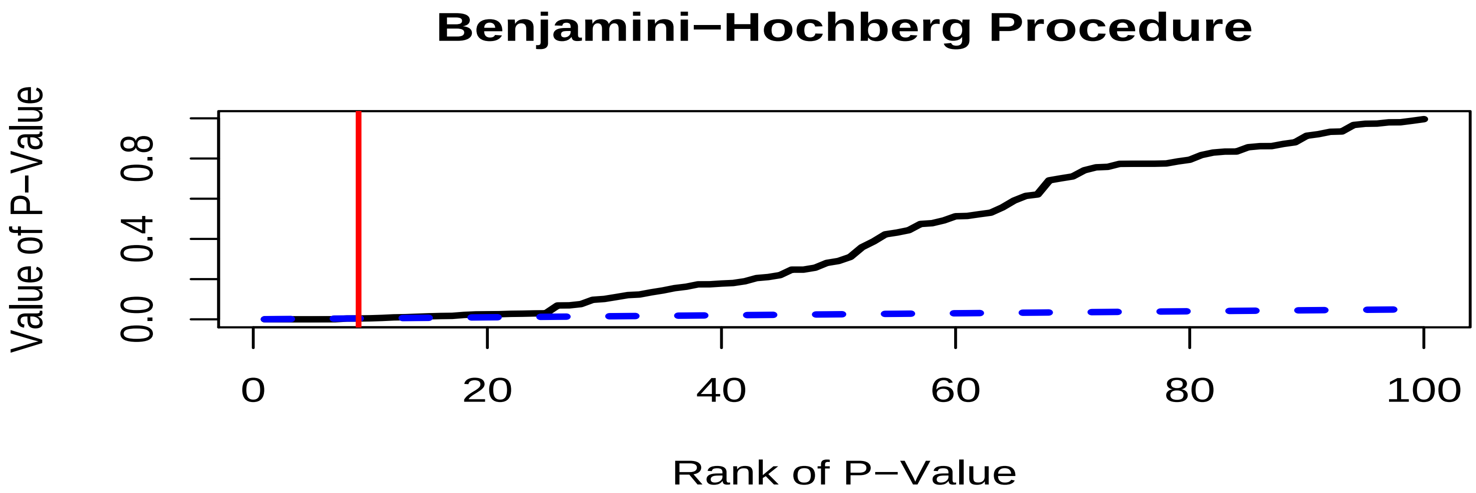
- Data Generation
- $M = 100$
- One-group model; by sufficiency, $n = 1$
- $X_m \sim N(\mu_m, 1), m = 1, 2, \ldots, M$
- $\mu_m = \xi_m I\{\theta_m = 1\}$
- $\xi_1, \xi_2, \ldots, \xi_M \sim |N(2, 1)|$
- $\theta_1, \theta_2, \ldots, \theta_M \sim Ber(p = .30)$
- $H_{m0} : \mu_m = 0$ versus $H_{m1} : \mu_m > 0$
- Tests: $\delta_m(X_m; \eta_m) = I\{X_m > \Phi^{-1}(1 - \eta_m)\}$
- Effect Size: $\gamma_m = \xi_m$

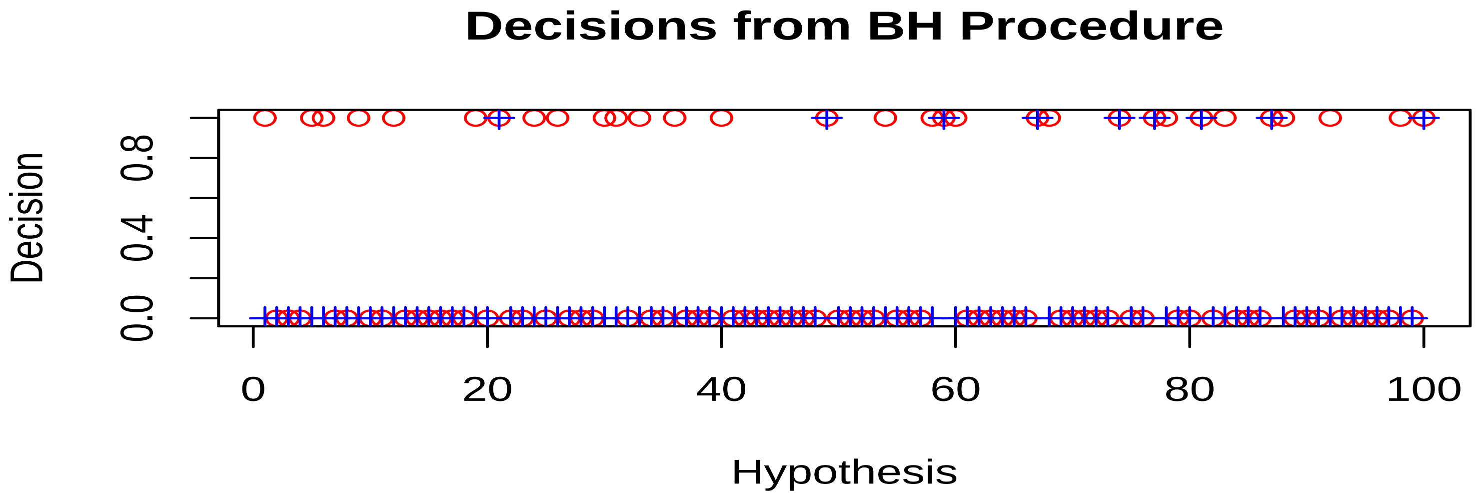
# Generalized BH Procedure



# While BH on Same Data



# Performances: GenBH and BH

- Interestingly, for the generated simulated data, both procedures have the same observed FDR and MDR.

| | Hypotheses Accepted | Hypotheses Rejected | Total |
|---|---|---|---|
| Correct Nulls | 69 | 0 | 69 |
| False Nulls | 22 | 9 | 31 |
| Total | 91 | 9 | 100 |

- Observed FDR = 0.
- Observed MDR = 70.9%.

# Simulation (100 reps): Error Rates (in %)

- First 10 replicates:

```
   GenBHfdr GenBHmdr     BHfdr    BHmdr
1  8.333333 60.71429 15.384615 60.71429
2  0.000000 67.64706  0.000000 70.58824
3  7.142857 51.85185  7.142857 51.85185
4  0.000000 82.75862  0.000000 82.75862
5  0.000000 58.82353  0.000000 64.70588
6  9.090909 64.28571  9.090909 64.28571
7  0.000000 55.17241  0.000000 62.06897
8  0.000000 54.54545  0.000000 57.57576
9  0.000000 63.88889  0.000000 72.22222
10 0.000000 66.66667  0.000000 83.33333
```

- Means (based on 100 replications):

| GenBHfdr | GenBHmdr | BHfdr | BHmdr |
|---|---|---|---|
| 3.450762 | 61.93329 | 4.134463 | 63.46749 |

# Concluding Remarks

- Needed: More analytical and simulated examination of properties of generalized BH.
- Power functions of individual tests *do* matter! Heeded a lesson of Neyman and Pearson.
- Invest your size on tests with neither too small nor too high a power.
- FWER-controlling procedure: anchor to developing FDR-controlling procedures.
- BUT, procedures probably *not* yet the truly optimal ones, since we started with $\delta_m$ that depended only on $(X_m, U_m)$.
- Is the Route to Real Optimality the Bayesian Way!? Currently being explored.