# MP Tests, $p$-Values, FWER and FDR Control in Multiple Testing


**Edsel A. Peña**

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

Department of Statistics, University of South Carolina

Columbia, SC 29208

Joint work with **Josh Habiger** and **Alex McLain**



Research Support from US NIH Grants


Presented July 2008 at the IBS Conference, Dublin, Ireland

# Multiple Testing

- Multiple hypotheses testing problems (MHTP) arise in many situations, notably in *large* $M$*, small* $n$ settings that characterize data sets dealing with multiple comparisons, microarrays, proteomics, and in other areas.
- Many MHTP procedures starts with the $p$-values of the individual tests associated with the $M$ pairs of null and alternative hypotheses.
- The famous Benjamini-Hochberg FDR-controlling procedure is an example of such a procedure.
- A very active area of research with many challenges. Proofs, e.g., in BH procedure, are quite challenging and interesting to understand!

# Features

- Many talks, and experts, on this area here at IBC ’08!
- Of utmost importance to scientists, e.g., medical researchers and biologists.
- Several recent publications, e.g., a recent issue of *Annals of Stat* has several, notably Efron’s paper.
- New books on this area, e.g., Dudoit and van der Laan (2007) in the Springer desk.
- Data for each of the pairs of hypotheses could come in complicated structures, e.g., Discrete, Continuous; ANOVA-type data; Regression-type; others.
- For moderate $M$, data could even be of the recurrent event type and/or longitudinal data type.

# Our Motivating Questions

- What is the *role* of the power functions of the individual tests in MHTP procedures?
- Are we exploiting the potential differences in their powers in current MHTP procedures?
- It appears that FWER-controlling procedures, such as the Sidak procedure, or FDR-controlling procedures, such as the BH procedure, assumes the same powers for each of the $M$ tests as the $p$-values are treated in a symmetric fashion.
- It seems rather unlikely, however, that the $M$ tests, especially with large $M$, would all have the same powers.

# A Look into History

- In 1920-30s, Neyman and Pearson recognized that in the search for optimal hypothesis tests, one must consider the alternative hypothesis. This is in contrast to the then-existing significance testing ($p$-value) approach.
- Led to development of Neyman-Pearson framework, resulting in the theory of most powerful (MP) and uniformly most powerful (UMP) tests, and exploitation of the monotone likelihood ratio (MLR) property.
- In MHTP we may view the configurations of the $M$ pairs of hypotheses as the 'alternative.' From the Neyman-Pearson lesson we should exploit this alternative configuration together with the individual powers of the tests to develop good procedures.

# Usual 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

- $\theta_m = I\{H_{m1}$ 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 $\delta_m : \mathcal{X}_m \to \{0,1\}$ 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.

# 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):

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

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.

# (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*.

# 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.

# Some Remarks

- In BH procedure, independence is only needed among those $X_m$s that correspond to true null hypotheses.
- In both Sidak and BH MHTPDFs, powers of the individual tests were not used in the procedures.
- Not clear if differences in powers of the individual tests are actually taken into account. Since $p$-values are treated in a symmetric fashion, it appears that differences in powers are not invoked.
- We now look into the potential effects of differences in powers. As in the Neyman-Pearson development, at each $m$, we first consider a simple null and a simple alternative.

# 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 $\eta_m$.

# Neyman-Pearson 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:

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

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

# FWER and MDR in Setting

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 and Uniqueness: PHM(08)

- *Theorem:* For any $\alpha \in (0,1)$, there always exists a size vector $(\eta_1(\alpha), \eta_2(\alpha), \ldots, \eta_M(\alpha))$ that solves the constrained optimization problem, hence an optimal MHTPDF that controls the FWER among the (restricted) class of decision functions considered.

- *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.

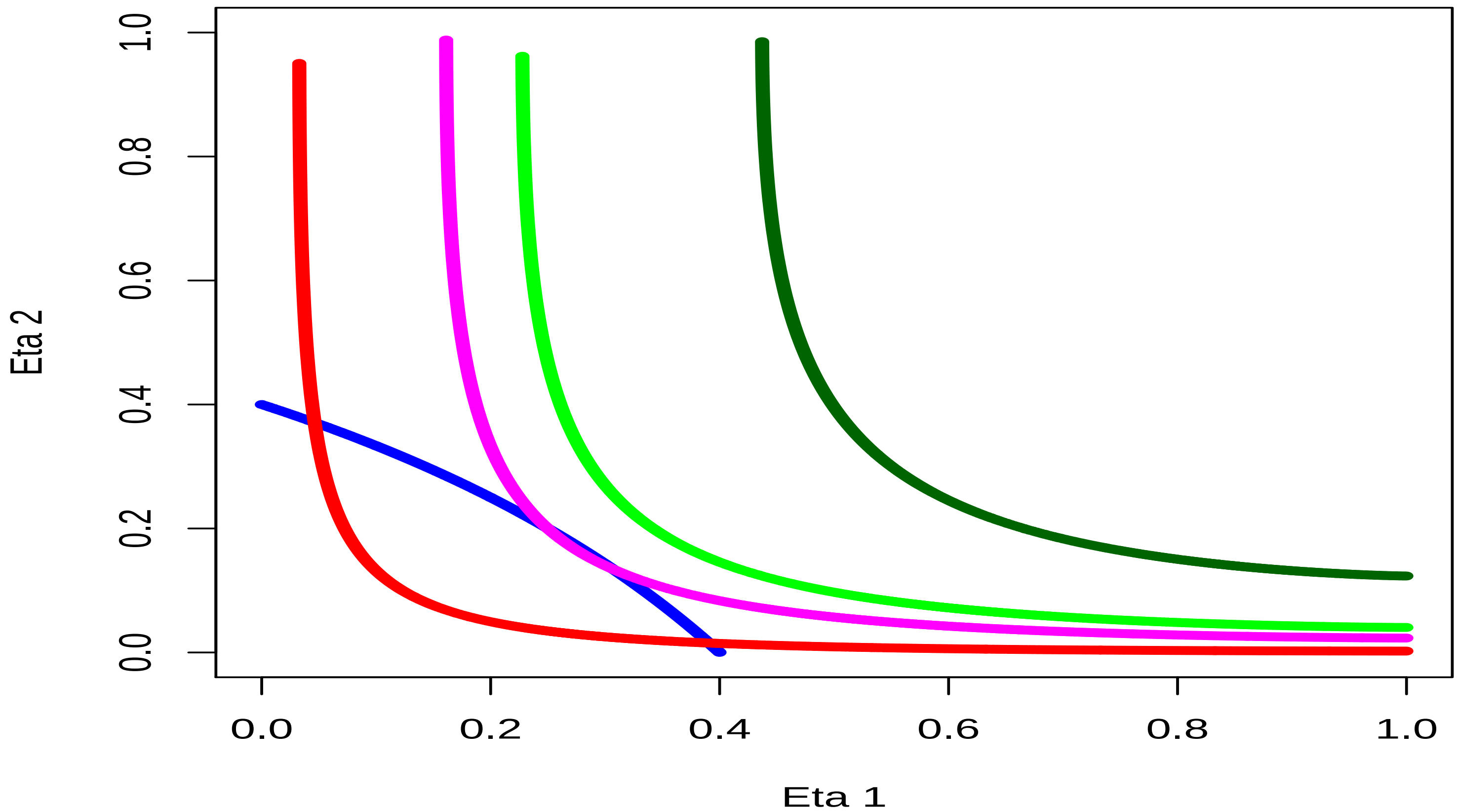
- *Corollary:* The Sidak MHTFDF obtains when the power functions $\eta_m \mapsto \pi_m(\eta_m)$ for $m = 1, 2, \ldots, M$ are identical.

# 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 $b$.



# 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

$$\pi'_m(\eta_m)(1-\eta_m) = \lambda \in \Re;$$

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

- 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 Properties

- 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 $H_{m0} : \mu_m \leq 0$ 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

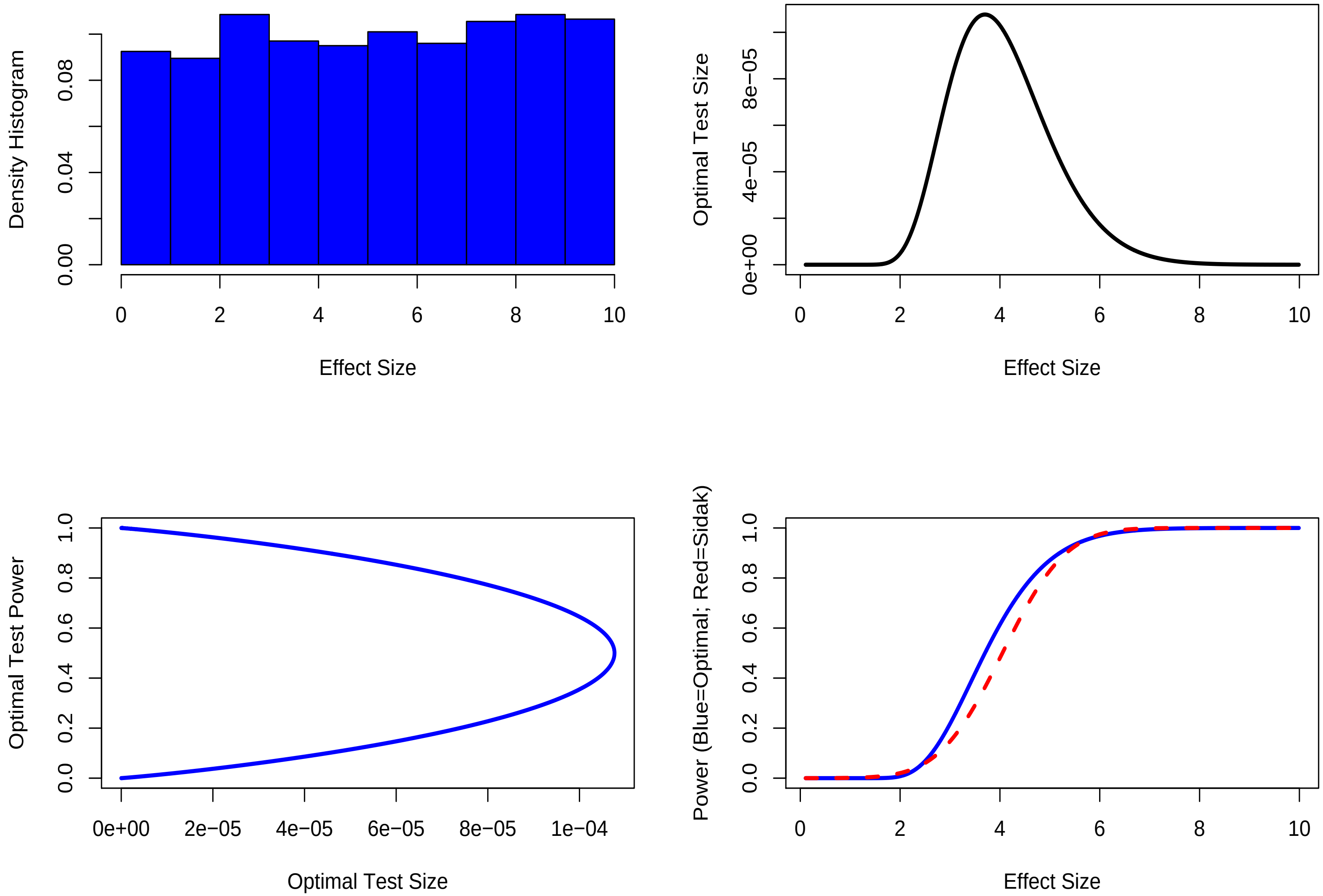
$$\pi_m(\eta_m) = 1 - \Phi(\Phi^{-1}(1 - \eta_m) - \gamma_m).$$

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

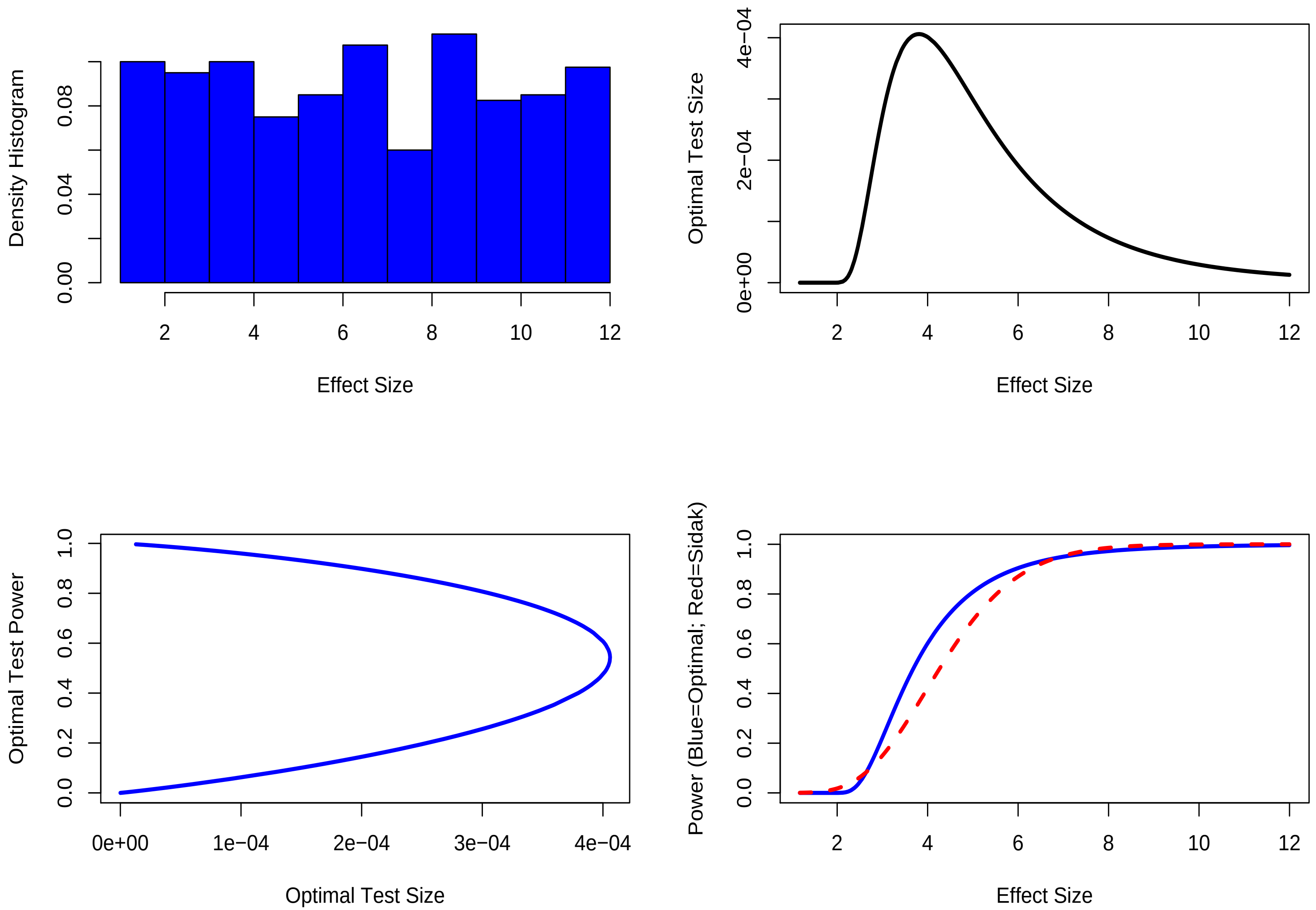
# Normal Example: Small $M$

| 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:** $M = 2000$; $\gamma_m \overset{IID}{\sim} U[.1, 10]$



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



# Some Observations

- Both the normal and exponential settings allowed the Lagrange solution approach.
- Also did an example with binomial distributions; computations more elaborate since this does not allow the Lagrange 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.
- Improvement in overall discovery rate over the Sidak procedure.

# A 'Size Investment' Lesson!

- An interesting result from the plots of the optimal size vectors is that small optimal sizes are allocated to those where the effect size is either small (which converts to low power) or the effect size is large (which converts to high power).
- Result is intuitive, in hindsight, and is indeed a simplistic investment strategy, albeit with respect to the allocation of test sizes:
- 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

- Data: $(\mathbf{X}, \mathbf{U}) = ((X_m, U_m), m = 1, 2 \ldots, M)$.
- Desired FDR-level: $q^*$. Define $\alpha_M^* \equiv \alpha_M^*(\mathbf{X}, \mathbf{U})$ via

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

- The proposed FDR-controlling MHTPDF is

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

- **Theorem:** Whatever the true $\theta_0$ is, $R_1(\delta^*(\alpha_M^*), \theta_0) \leq q^*$.
- Formal Proof: In a forthcoming manuscript.

# Intuition & Motivation (Informal Proof)

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

$$E\left\{\sum_m \delta_m(\eta_m(\alpha)(1-\theta_m)\right\} \leq M[1-\prod_m(1-\eta_m(\alpha)]] = M\alpha$$

$$Q_M(\delta^*(\alpha)) \ \tilde{\leq} \ \frac{M\alpha}{\sum_m \delta_m(\eta_m(\alpha))}$$

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

# Concluding Remarks

- Power functions (as functions of their size) of individual tests matter! Heeded an old lesson of Neyman and Pearson.
- Invest your size on tests with neither too small nor too high a power.
- FWER-controlling procedure an anchor to developing better FDR-controlling procedures.
- However, most probably, the procedures are not yet *the* truly optimal ones, since we started with a test $\delta_m$ that depended only on the data $(X_m, U_m)$.
- Route to Real Optimality? Is the Bayesian Way!?