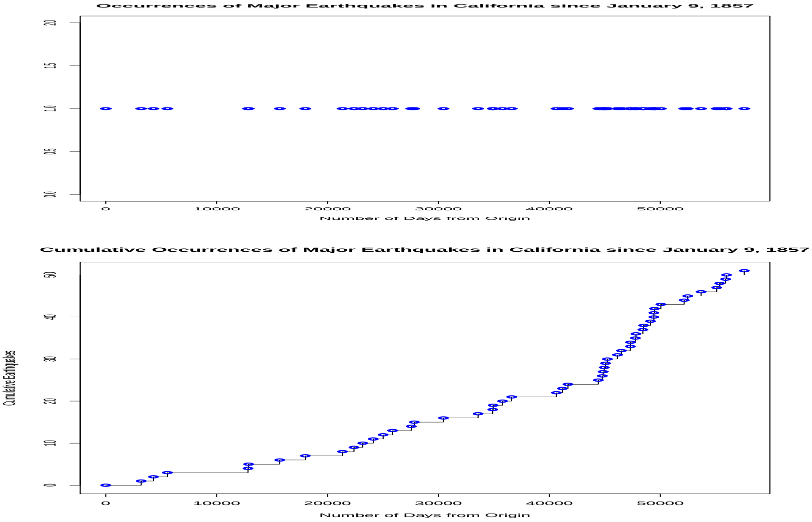
On Multiple Decision-Making

Edsel A. Peña

Department of Statistics
University of South Carolina
Columbia, SC 29208

SAMSI Workshop on
Games and Decisions in Risk and Reliability (GDRR)
May 18, 2016

Major Earthquakes in California, 1857-2015

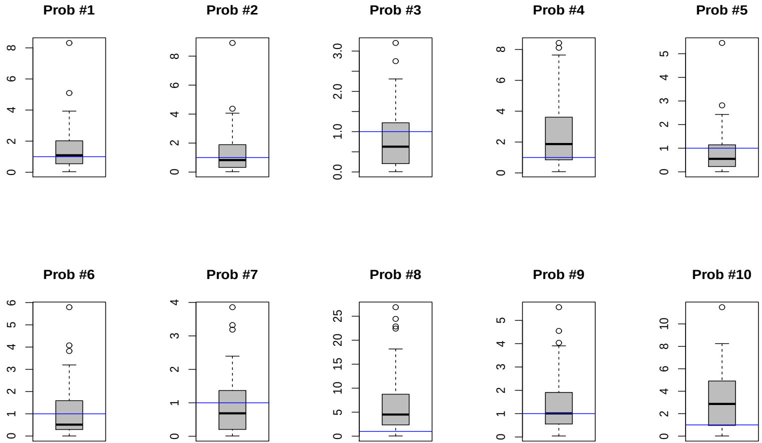


Questions: From an HPP(λ)? If so, is $MTBE = \lambda^{-1} > 1000$ days?

Multiple Testing Problem

$M = 10$ HPPs

Which of the 10 HPPs have $\lambda < 1$ Based on the $n = 50$ Observations?



Individual Decision Functions

- Decision Problems: $m \in \{1, 2, \ldots, M\}$
- Model (since HPP):

$$X_{m1}, X_{m2}, \ldots, X_{mn} \quad IID \quad Exp(\lambda_m)$$

- Hypotheses:

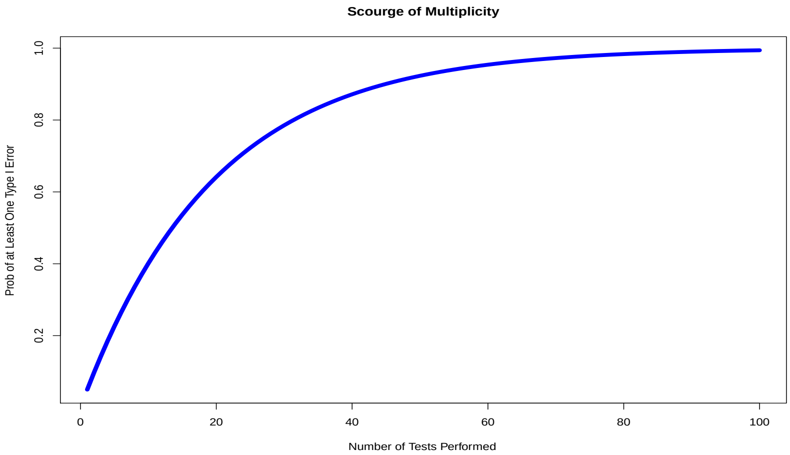
$$H_{m0} : \lambda_m \geq 1 \quad versus \quad H_{m1} : \lambda_m < 1$$

- Level of Significance: α, e.g., 0.05.
- Decision Function:

$$\delta_m(\mathbf{x}_m) = I\left\{2\sum_{i=1}^{n} X_{mi} > \chi^2_{2n;\alpha}\right\}$$

That is, decide for H_{m1} if $2S_m = 2\sum_{i=1}^{n} X_{mi} > \chi^2_{2n;\alpha}$.

Perils of Multiple Decision-Making: Fixed α



Relevance of Multiple Decisions

- Post-hoc analysis after a significant ANOVA test.
- Variable selection in regression problems.
- Genetic and/or proteomics studies.
- Engineering and reliability settings.
- Business and/or financial (investment) situations.
- University admission of students.
- Sporting decisions.
- Examples in yesterday's talks.

The General Multiple Decision Problem

- To discover the value (truth) of a parameter vector

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

- $\theta_m = 1$ means the mth alternative hypothesis H_{m1} is true.
- $\theta_m = 0$ means the mth null hypothesis H_{m0} hypothesis is true.
- Decision Maker's Goal: To choose an action vector

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

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

False and Missed Discoveries

- False Discovery [Type I Error]: action $a_m = 1$ when $\theta_m = 0$.

$$a_m(1 - \theta_m) = 1$$

- Missed Discovery [Type II Error]: action $a_m = 0$ when $\theta_m = 1$.

$$(1 - a_m)\theta_m = 1$$

- Practical Consequences of Errors of Decision:
  - False Discovery: Ends up *Chasing Noise as Signal!*
  - Missed Discovery: *Overlooking Key to Solving Problem.*
  - Reported results **NOT** Reproducible.
  - Non-Reproducible Results: Crisis in Science. [John Ionnadis]

Globally Assessing Multiple Actions: Losses

- Family-wise error indicator (FWEI):

$$L_0(a,\theta) = I\left\{\sum_{m=1}^{M} a_m(1-\theta_m) > 0\right\} = \text{At least one Type I Error?}$$

- False Discovery Proportion (FDP):

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

- Missed Discovery Proportion (MDP):

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

Making Multiple Decisions from Data, Possibly BIG

- Obtain data (e.g., microarrays, brain data, Netflix, etc.):

$$X \in \mathfrak{X} = \text{Data Space}$$

- X maybe an $M \times n$ matrix; could be more complex.
- Probabilistic Structure:

$$X \sim P, \quad P \text{ is a Joint Probability Distribution}$$

- Marginal Components (Data and Distribution):

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

- Marginal Parameters: $\theta_m = \theta_m(P_m)$
- An Example:

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

Multiple Decision Functions

- Multiple Decision Function:

$$\delta : \mathfrak{X} \text{ (Data Space) } \to \mathfrak{A} = \{0,1\}^M \text{ (Action Space)}$$

- $\delta = (\delta_1, \delta_2, \ldots, \delta_M)$ with $\delta_m : \mathfrak{X} \to \{0,1\}$
- Remark: δ_m may use the whole data, not just X_m.
- $\mathfrak{D}$: 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):

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

- Goal: Choose $\delta \in \mathfrak{D}$ with small risks, whatever P is.

Weak and Strong FWER or FDR Control by an MDF δ

- Weak FWER or FDR Control at Level q: For any P_0 with $\theta_m(P_0) = 0$ for all m,

$$R_0(\delta, P_0) \leq q \quad \text{or} \quad R_1(\delta, P_0) \leq q.$$

- Strong FWER or FDR Control at Level q: For *any* P,

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

- Question: Could we obtain classes of MDFs strongly controlling the FWER or FDR?
- Game-theoretic aspects: 'Nature' vs cooperating multiple 'players'.

Revisiting Single Decision Problem

- Two Population Distributions:

$$F = N(\mu_1, \sigma^2), \text{ Treated}; \quad G = N(\mu_0, \sigma^2), \text{ Control}$$

- Decision Problem is to choose between

$$H_0 : \mu_1 = \mu_0 \Leftrightarrow \theta = 0$$
$$H_1 : \mu_1 > \mu_0 \Leftrightarrow \theta = 1$$

- H_0 usually STATUS QUO; H_1 usually RESEARCH hypothesis.
- Decision based on random samples:

$$\mathbf{X} = (X_1, X_2, \ldots, X_m) \sim F; \quad \mathbf{Y} = (Y_1, Y_2, \ldots, Y_n) \sim G$$

- Decision Function: $(\mathbf{x}, \mathbf{y}) \mapsto \delta(\mathbf{x}, \mathbf{y}) \in [0, 1]$
- $\delta(\mathbf{x}, \mathbf{y})$ is probability of deciding for H_1.

T-Test Procedure

- T-Test Statistic:

$$T(\mathbf{X}, \mathbf{Y}) = \frac{\bar{X} - \bar{Y}}{S_p\sqrt{\frac{1}{m} + \frac{1}{n}}}; \quad S_P^2 = \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}$$

- Decision Function of Size α:

$$\delta(\mathbf{x}, \mathbf{y}; \alpha) = I\left\{T(\mathbf{x}, \mathbf{y}) > \mathcal{T}^{-1}(1-\alpha, m+n-2)\right\}$$

- Alternative Approach: P-Value Statistic is

$$P(\mathbf{x}, \mathbf{y}) = \Pr\{T(\mathbf{X}, \mathbf{Y}) > T(\mathbf{x}, \mathbf{y}) | H_0\}$$

- Equivalent Decision Function:

$$\delta(\mathbf{x}, \mathbf{y}; \alpha) = I\{P(\mathbf{x}, \mathbf{y}) < \alpha\}$$

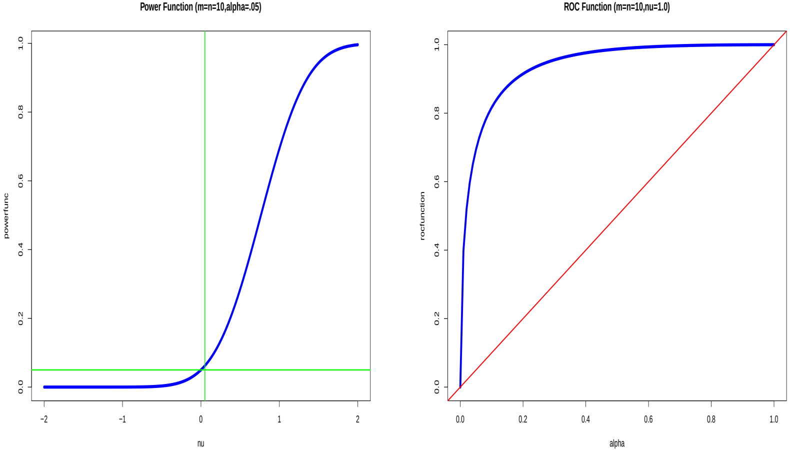
Power Function and ROC Function

- (Standardized) Effect Size: $\nu = (\mu_1 - \mu_0)/\sigma$
- An Important Function:

$$\pi(\nu;\alpha) = \Pr\{\delta(\mathbf{X},\mathbf{Y}) = 1 | \text{Effect Size is } \nu; \text{ Level } \alpha\}$$
$$= 1 - \mathcal{T}\left[\mathcal{T}^{-1}(1-\alpha, m+n-2), m+n-2, \sqrt{\frac{mn}{m+n}}\nu\right]$$

- Power Function: $\pi(\nu;\alpha)$ when viewed as function of effect size ν with level α fixed.
- ROC Function: $\pi(\nu;\alpha)$ when viewed as function of level α with effect size ν fixed.
- Sample Size Determination: for fixed ν, α, and value of power, solve for (m, n).

Plots of Power and ROC Functions



Decision Process and Multiple Decision Process

- In the two-sample T-test example, decision function depends on $\alpha \in [0,1]$. Leads to the notion of a decision process:

$$\Delta = (\delta(\mathbf{x}; \alpha) \equiv \delta(\alpha) : \alpha \in [0,1])$$

- For multiple decision problem: Multiple Decision Process:

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

- Decision Process for mth Component:

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

- Usual Approach: Pick a decision function δ_m from Δ_m using the same α.
- Common Choices for α: (weak) FWER Threshold of q use:

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

Notion of Size Function (Size-Picker)

- A size function is an

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

continuous, strictly increasing, $A(0) = 0$, and $A(1) \leq 1$.

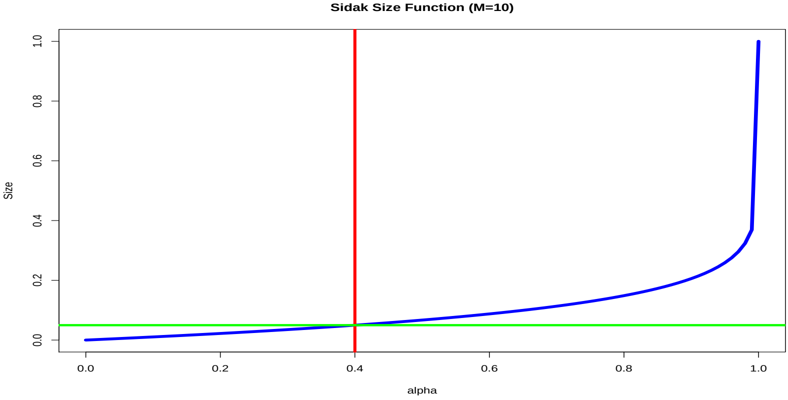
- Bonferroni size function: $A(\alpha) = \alpha/M$
- Sidak size function: $A(\alpha) = 1 - (1-\alpha)^{1/M}$

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

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

Illustration of Size Function

$$\text{Sidak: } A(\alpha) = 1 - (1-\alpha)^{1/10}$$



Multiple Decision Size Function

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

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

- **Global Size Condition:**

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

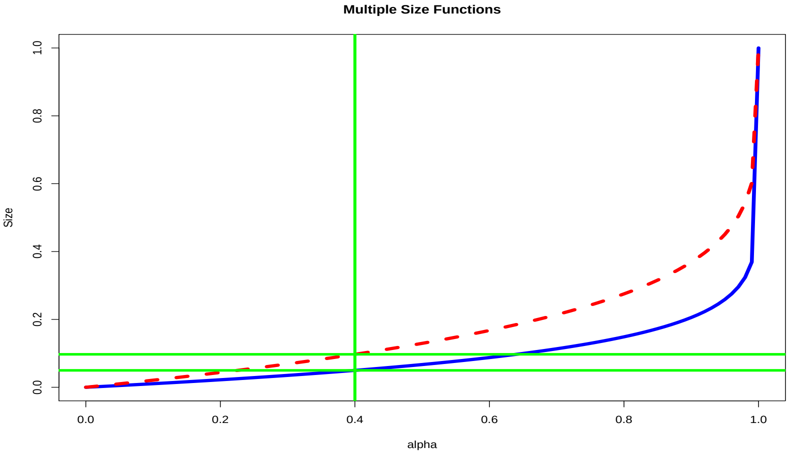
- Given a $\Delta = (\Delta_m : m \in \mathcal{M})$ and an $\mathbf{A} = (A_m : m \in \mathcal{M})$, multiple decision function is chosen according to

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

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

$$R_0(\delta(\alpha), P) \quad = \quad 1 - \prod_{m \in \mathcal{M}} [1 - A_m(\alpha)] = \alpha$$

Illustration of Multiple Size Functions



Towards Strong FWER Control

Given a MDP $\Delta = (\Delta_m)$ and MDS $\mathbf{A} = (A_m)$, the MDF at α given by $\delta = (\delta_m[A_m(\alpha)], m \in \mathcal{M})$ has FWER:

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

Question: Given a FWER level q, what is the best α?

Nostalgia (for our Spanish Friends): **Paul, the Oracle!**

'Best' Choice of α

- An Oracle's Choice:

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

- But, P is unknown to us, hence $\theta_m(P)$ and $\alpha^{\dagger}(q;P)$ are unknown to us.
- However, we could estimate $\theta_m(P)$ by

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

- The Oracle's choice is then estimated by

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

Strong FWER-Controlling MDF

- Chosen Multiple Decision Function:

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

- Theorem (Peña, Habiger, Wu, 2011, Ann Stat; 2015, Metrika)
  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. Thus, it is an MDF achieving strong FWER control at level q.

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

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

- Observe:

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

FDR-Controlling MDFs

- Let

$$\alpha^*(q) = \sup\left\{\alpha \in [0,1] : \sum_{m=1}^{M} A_m(\alpha) \le q \sum_{m=1}^{M} \delta_m[A_m(\alpha)]\right\}$$

- Chosen Multiple Decision Function:

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

- Theorem (Peña, et al, 2011, Ann Stat; 2015, Metrika)

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

$$R_1(\delta^*(q), P) \le q,$$

whatever P *is. [Remark: Needed is a Noether-type condition on* $\mathbf{A}$*.]*

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 (BH) 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.

Improving on the BH FDR-Controlling MDF

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

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

- 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.

Case with Simple Nulls and Simple Alternatives

- For each $m = 1, 2, \ldots, M$, $X_m \sim f_m$ and to decide between

$$H_{m0} : f_m = f_{m0} \Leftrightarrow \theta_m = 0$$

$$H_{m1} : f_m = f_{m1} \Leftrightarrow \theta_m = 1$$

- Neyman-Pearson MP Decision Process for each m has, with $L_m(x_m) = f_{m1}(x_m)/f_{m0}(x_m)$:

$$\delta_m(x_m; \alpha) = \begin{cases} 1 & \text{if} \quad L_m(x_m) > c_m(\alpha) \\ \gamma_m(\alpha) & \text{if} \quad L_m(x_m) = c_m(\alpha) \\ 0 & \text{if} \quad L_m(x_m) < c_m(\alpha) \end{cases}.$$

- ROC Functions:

$$\alpha \mapsto \pi_m(\alpha) \equiv \pi_m(\alpha; (f_{m0}, f_{m1}))$$

- Concave, continuous, increasing, and twice-differentiable.

Optimal Size Functions

Theorem (Peña, et al, 2011, Ann Stat & Metrika)

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

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

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

Equivalent to

$$\left[\mathbf{I} - \frac{1}{M}\mathbf{J}\right] \kappa(\mathbf{z}) + \mathbf{l}(\mathbf{z}) = \mathbf{1}\left[\frac{1}{M}\log(1-\alpha)\right];$$

$\mathbf{z} = (A_m(\alpha), m = 1:M)^{\mathrm{t}}$, $\kappa(\mathbf{z}) = (\log[\pi'_m(z_m)], m = 1:M)^{\mathrm{t}}$, *and* $\mathbf{l}(\mathbf{z}) = (\log(1 - z_m), m = 1:M)^{\mathrm{t}}$.

With f_{m0} and f_{m1} Normal Distributions

- $X_m \sim N(\mu_m, \sigma = 1), m = 1, 2, \ldots, M$
- $H_{m0} : \mu_m = 0$ versus $H_{m1} : \mu_m = \gamma_m$
- For each m, Neyman-Pearson MP decision process.

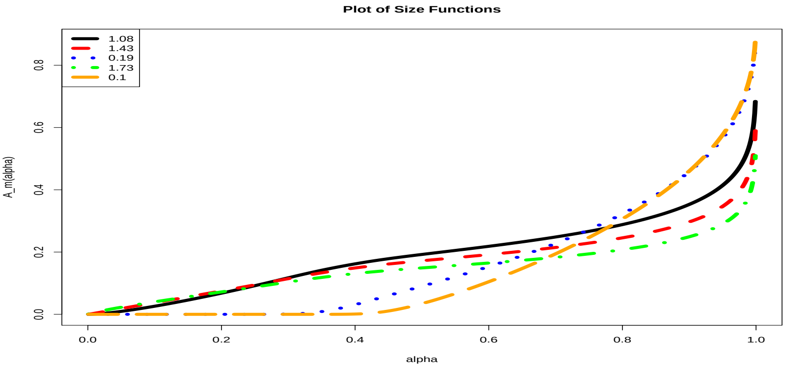
$$\Delta_m = (\delta_m(\alpha) : \alpha \in [0, 1])$$

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

- ROC Function for the *m*th NP MP Decision Process:

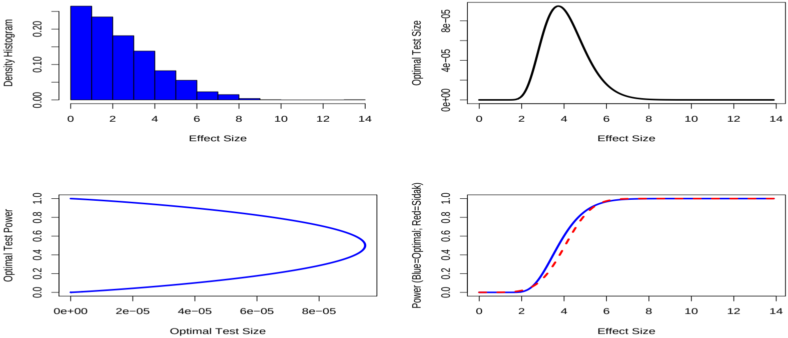
$$\pi_m(\alpha) = 1 - \Phi\left[\Phi^{-1}(1 - \alpha) - \gamma_m\right]$$

Shapes of Optimal Multiple Size Function: $M = 5$



Optimal Test Sizes vs Effect Sizes: $M = 2000$

Effect Sizes Generated Via: $\gamma_m \stackrel{IID}{\sim} |N(0,3)|$



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)!
- Concentrate on those where it is a bit uncertain, since your **differential gain** in overall discovery rate would be greater!
- On Some Interesting Consequences.
  - *On Graduate Student Advising.*
  - *US Presidential Election Strategy.*
  - Generally: Allocating Resources.

Simulations: Comparing BH with $\delta^*(q)$

Normally-distributed data; $q^* = 10\%$; $M = 100$; 1000 Replications

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

With Exponentially-Distributed Data; 1000 reps; $p = \#$ of correct alternatives; $q = 20\%$; Effect Sizes $\xi_m \sim U[.25, .75]$

M	$n_m = n$	p	BH		$\delta^*(q)$	
			FDR	MDR	FDR	MDR
100	5	.2	15.93574	55.56008	16.03789	55.24426
100	10	.2	15.78390	38.48281	14.97451	38.12848
100	30	.2	16.09659	10.88535	14.93114	8.026157
100	5	.3	13.42253	46.06943	13.24371	45.90983
100	10	.3	13.76047	31.97512	13.65171	30.47509
100	30	.3	13.80077	5.639566	13.89979	2.626881
100	5	.5	10.37083	41.65425	10.06751	41.40670
100	10	.5	10.11820	26.34404	10.11692	23.57873
100	30	.5	10.03628	7.615538	9.590189	4.151696

Same Setting but Only Four Values of Effect Sizes

Effect Sizes Used

$$m = 1 : 25 : \ \xi_m = .1; \ m = 26 : 50 : \ \xi_m = .3$$

$$m = 51 : 75 : \ \xi_m = .5; \ m = 76 : 100 : \ \xi_m = .7$$

M	$n_m = n$	p	BH		$\delta^*(q)$	
			FDR	MDR	FDR	MDR
100	5	.2	16.3581	39.05525	15.79368	36.95671
100	10	.2	15.86433	26.73459	16.05328	22.97558
100	30	.2	16.0207	9.896752	15.76027	5.316628

Acknowledgements

- This talk based on two papers in AoS (2011) and Metrika (2015) co-authored with former students
- Josh Habiger, Oklahoma State Univ.
- Wensong Wu, Florida International Univ.
- NSF and NIH research grants.
- Organizers of this Workshop: Refik Soyer & Fabrizio Ruggeri & SAMSI.