Lec 4: Testing Multiple Hypotheses

Xin Xing

The Multiple Testing Problem in Sparse Linear Models

- Model Setup:

$$Y = X\beta + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I_n)$$

  - $X \in \mathbb{R}^{n \times p}$: design matrix
  - $\beta \in \mathbb{R}^p$: sparse coefficient vector
  - Most $\beta_j = 0$ (sparsity)
- Testing Problem:

$$H_{0j} : \beta_j = 0 \text{ vs } H_{1j} : \beta_j \neq 0$$

$$\text{for } j = 1, \ldots, p$$

- Challenges:
  - High-dimensional: $p \gg n$ possible
  - Many tests: risk of false discoveries
  - Dependencies between tests

Error Types in Multiple Testing

	H_0 True	H_0 False	Total
Reject H_0	V (FP)	S (TP)	R
Accept H_0	U (TN)	T (FN)	m-R
Total	m_0	$m - m_0$	m

- Key Metrics:
  - V (Number of False Positives)
  - FDP = V/R (False Discovery Proportion)
  - FDR = E[FDP] (False Discovery Rate)
  - Power = E[S/(m-m_0)](*TruePositiveRate*)

The Multiple Testing Problem

Examples:

- Test the safety of a drug in terms of a dozen different side effects
- Test whether a disease is related to 10,000 different gene expressions

Simulation Setup

- Sample size (n): 1000
- Number of variables (p): 100
- True non-zero coefficients (k): 10
- Design matrix: Orthogonal (via QR decomposition)
- Signal strength: Uniform$(0.5, 2) \times \{-1, 1\}$

Model Structure

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

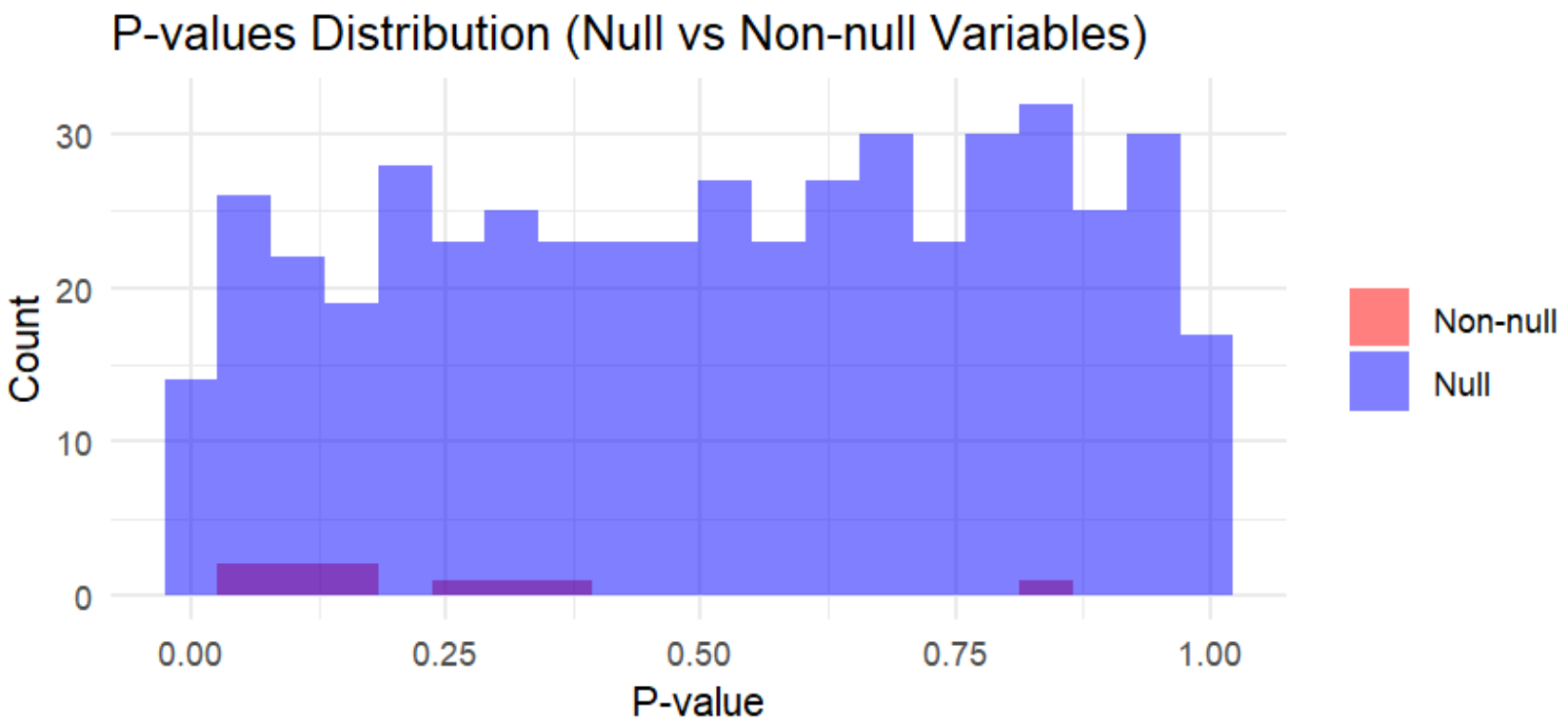
where:

$\mathbf{X}$: $n \times p$ orthogonal design matrix

$\boldsymbol{\beta}$: p-dimensional sparse vector (10% non-zero)

$\boldsymbol{\epsilon}$: Gaussian noise $\mathcal{N}(0, 1)$

P-value Distribution



P-values

1: 0.000354 2: 0.001940 3: 0.010713 4: 0.017502 5: 0.019476
6: 0.024788 7: 0.038100 8: 0.067764 9: 0.076740 10: 0.087196
11: 0.099932 12: 0.128348 13: 0.138694 14: 0.160054 15: 0.166845
16: 0.181025 17: 0.181632 18: 0.185423 19: 0.186156 20: 0.212532

Family-wise Error Rate (FWER)

The family-wise error rate (FWER) is the probability of falsely rejecting at least one true null hypothesis:

$$\mathbb{P}[V \geq 1]$$

A procedure controls FWER at level α if $\mathbb{P}[V \geq 1] \leq \alpha$, regardless of the (possibly unknown) number of true null hypotheses m_0.

The Bonferroni Correction

To control FWER at level α:

- Perform each test at significance level α/m, instead of level α
- Simple but can be conservative
- Without loss of generality, let $H_0^{(1)}, \ldots, H_0^{(m_0)}$ be the true null hypotheses

Verification:

$$\begin{aligned}
\mathbb{P}[V \geq 1] &= \mathbb{P}\left[\{\text{reject } H_0^{(1)}\} \cup \ldots \cup \{\text{reject } H_0^{(m_0)}\}\right] \\
&\leq \mathbb{P}\left[\text{reject } H_0^{(1)}\right] + \ldots + \mathbb{P}\left[\text{reject } H_0^{(m_0)}\right] \\
&= \frac{\alpha}{m} + \ldots + \frac{\alpha}{m} = \frac{\alpha m_0}{m} \leq \alpha
\end{aligned}$$

Error Types in Multiple Testing

	H_0 True	H_0 False	Total
Reject H_0	V (FP)	S (TP)	R
Accept H_0	U (TN)	T (FN)	m-R
Total	m_0	$m - m_0$	m

- Key Metrics:
  - FDP = V/R (False Discovery Proportion)
  - FDR = E[FDP] (False Discovery Rate)
  - Power = E[S/(m-m_0)](*TruePositiveRate*)

False Discovery Rate (FDR)

Controlling FWER $\mathbb{P}[V \geq 1]$ may be too conservative and greatly reduce our power to detect real effects, especially when m is large.
Instead, consider the false discovery proportion (FDP):

$$\text{FDP} = \begin{cases} V/R & R \geq 1 \\ 0 & R = 0 \end{cases}$$

The false discovery rate (FDR) is the expected value $\mathbb{E}[\text{FDP}]$.

Estimate FPs

- Let $R(t)$ be the number of p-values $\leq t$ for $t \in (0, 1)$
- When rejecting hypotheses with p-value $\leq t$:
  - Expect to falsely reject tm_0 null hypotheses
  - (null p-values follow Uniform(0,1))
- Estimated false discovery proportion: $\frac{tm_0}{R(t)}$
- Conservative upper bound (since m_0 unknown): $\frac{tm}{R(t)}$

rank 1: 0.0354
rank 2: 0.0970
rank 3: 0.3571
rank 4: 0.4375
rank 5: 0.3895
rank 6: 0.4131
rank 7: 0.5443
rank 8: 0.8470
rank 9: 0.8527
rank 10: 0.8720

Motivating the BH Procedure

- Let $R(t)$ be the number of p-values $\leq t$ for $t \in (0, 1)$
- For $t = P_{(r)}$ (the rth smallest p-value):

$$\frac{tm}{R(t)} \leq q \iff P_{(r)} \leq \frac{qr}{m}$$

- BH procedure: Choose t data-dependently to maximize rejections while maintaining $\frac{tm}{R(t)} \leq q$

The Benjamini-Hochberg Procedure

To control FDR at level q:

1. Sort the p-values. Call them $P_{(1)} \leq \ldots \leq P_{(m)}$
2. Find the largest r such that $P_{(r)} \leq \frac{qr}{m}$
3. Reject the null hypotheses $H_{(1)}, \ldots, H_{(r)}$

Theorem (Benjamini and Hochberg, 1995):
If the test statistics are independent, then the FDR of this procedure satisfies:

$$\text{FDR} \leq \frac{n_0 q}{n} \leq q$$

BH FDR Control: Optional Stopping Approach

- Setup:
  - $H_1, \ldots, H_m$ hypotheses with p-values $p_1, \ldots, p_m$
  - I_0: indices of true nulls, $m_0 = |I_0|$
  - $R = \max\{k : p_{(k)} \leq \frac{qk}{m}\}$

Constructing the Process

- For $t \in [0, 1]$, define:

$$V(t) = \sum_{i \in I_0} 1\{p_i \leq t\}$$

- Key observation:

$$\{V(t), t \in [0, 1]\} \text{ is a process in continuous time}$$

- Let $\mathcal{F}_t = \sigma(1\{p_i \leq s\} : s \leq t, i \in I_0)$

The Process V(t)/t

- BH analogy:
  - Instead of money, track proportion of false discoveries
  - Instead of time n, use threshold t
  - $\frac{V(t)}{t}$ is like "fortune per unit time"
- Process Construction:

$$\text{Let } Y_t = \frac{V(t)}{t} \text{ for } t \in (0, 1]$$

(proportion of false discoveries among rejections)

- **Key Assumption:**

$$\frac{V(s)}{s} \leq m_0 \text{ for all } s \in (0, 1]$$

(enforced by BH stopping rule)

Setup: The Martingale M_t

- Define:

$$M_t = V(t) - m_0 t$$

where:

  - $V(t)$ = number of false discoveries (true nulls rejected) up to threshold t
  - m_0 = total number of true null hypotheses
  - $t \in (0, 1]$ = p-value threshold

- Goal: Show M_t is a martingale

$$E[M_t | \mathcal{F}_s] = M_s \text{ for } 0 < s < t \leq 1$$

Filtration $\mathcal{F}_s$

- $\mathcal{F}_s$ = information available at threshold s
  - Which hypotheses have p-values $\leq s$
  - How many of those are false discoveries: $V(s)$
  - How many total rejections: $R(s)$
- At threshold s:
  - We've observed $V(s)$ false discoveries
  - We don't know which specific nulls are true/false
  - But we know there are m_0 true nulls total

Key Insight: Independence

- Under true null H_i, p-value $P_i \sim \text{Uniform}(0, 1)$
- For any threshold interval $(s, t]$:

$$\Pr(P_i \in (s, t] \mid P_i > s) = \frac{t - s}{1 - s}$$

- Under independence assumption:

$$\Pr(P_i \leq t \mid P_i > s) = \frac{t - s}{1 - s}$$

- For small intervals or under PRDS:

$$\Pr(P_i \in (s, t]) \approx t - s$$

Computing $E[V(t)|\mathcal{F}_s]$: Step 1

- Decompose $V(t)$:

$$V(t) = V(s) + \underbrace{[V(t) - V(s)]}_{\text{new false discoveries in } (s,t]}$$

- At threshold s:
  - $V(s)$ is already observed (known)
  - $V(t) - V(s)$ counts additional false discoveries in $(s, t]$
- Therefore:

$$E[V(t)|\mathcal{F}_s] = V(s) + E[V(t) - V(s)|\mathcal{F}_s]$$

Computing $E[V(t)|\mathcal{F}_s]$: Step 2

- New false discoveries in $(s, t]$:

$$V(t) - V(s) = \sum_{i \in H_0} 1\{s < P_i \leq t\}$$

  where H_0 = set of true null hypotheses
- Conditional expectation:

$$E[V(t) - V(s)|\mathcal{F}_s] = \sum_{i \in H_0} E[1\{s < P_i \leq t\}|\mathcal{F}_s]$$

- Key observation: Given $\mathcal{F}_s$, we know which hypotheses have $P_i \leq s$

Computing $E[V(t)|\mathcal{F}_s]$: Step 3

- Split the sum:

$$\sum_{i \in H_0} E[1\{s < P_i \leq t\}|\mathcal{F}_s]$$

$$= \sum_{i \in H_0, P_i \leq s} E[1\{s < P_i \leq t\}|\mathcal{F}_s] + \sum_{i \in H_0, P_i > s} E[1\{s < P_i \leq t\}|\mathcal{F}_s]$$

- First sum $= 0$ (since $P_i \leq s$ implies $P_i \notin (s, t]$)
- Second sum: For i with $P_i > s$:

$$E[1\{s < P_i \leq t\}|\mathcal{F}_s] = \Pr(P_i \in (s, t]|P_i > s) = t - s$$

Computing $E[V(t)|\mathcal{F}_s]$: Step 4

- How many true nulls have $P_i > s$?
  - Total true nulls: m_0
  - True nulls already discovered (have $P_i \leq s$): $V(s)$
  - True nulls remaining: $m_0 - V(s)$
- Therefore:

$$E[V(t) - V(s)|\mathcal{F}_s] = (m_0 - V(s)) \cdot (t - s)$$

- Wait, this doesn't give us what we want!

Correction: Unconditional Probability

- Actually, under independence (or PRDS):

$$\Pr(s < P_i \leq t) = t - s \text{ for any true null } i$$

- For each of m_0 true nulls:

$$E[1\{s < P_i \leq t\}|\mathcal{F}_s] = t - s$$

(independence from past)

- Therefore:

$$E[V(t) - V(s)|\mathcal{F}_s] = m_0(t - s)$$

- This gives:

$$\boxed{E[V(t)|\mathcal{F}_s] = V(s) + m_0(t - s)}$$

Why M_t is a Martingale

- Now compute $E[M_t|\mathcal{F}_s]$:

$$E[M_t|\mathcal{F}_s] = E[V(t) - m_0 t|\mathcal{F}_s]$$

$$= E[V(t)|\mathcal{F}_s] - m_0 t$$

$$= [V(s) + m_0(t-s)] - m_0 t$$

$$= V(s) + m_0 t - m_0 s - m_0 t$$

$$= V(s) - m_0 s$$

$$= M_s$$

- Therefore, $M_t = V(t) - m_0 t$ is a **martingale**!

Intuition: Why Martingale?

- $V(t)$ = cumulative false discoveries up to threshold t
- Expected drift: Each true null adds $(t - s)$ probability

$$E[\Delta V] = m_0 \cdot (t - s) = m_0 \Delta t$$

- Subtracting this drift:

$$M_t = V(t) - m_0 t$$

removes the predictable trend

- Result: Zero expected change

$$E[M_t - M_s | \mathcal{F}_s] = 0$$

(classic martingale property)

Submartingale Property: Step 1

- For $Y_t = V(t)/t$:

$$E[Y_t|\mathcal{F}_s] = E\left[\frac{V(t)}{t}\middle|\mathcal{F}_s\right]$$

- Using linearity of conditional expectation:

$$= \frac{1}{t}E[V(t)|\mathcal{F}_s]$$

- Substitute $E[V(t)|\mathcal{F}_s] = V(s) + m_0(t-s)$:

$$= \frac{V(s) + m_0(t-s)}{t}$$

Submartingale Property: Step 2

- We need to show: $E[Y_t|\mathcal{F}_s] \geq Y_s$

$$\frac{V(s) + m_0(t-s)}{t} \geq \frac{V(s)}{s}$$

- Multiply both sides by $st > 0$:

$$s[V(s) + m_0(t-s)] \geq t \cdot V(s)$$

- Expand:

$$sV(s) + sm_0(t-s) \geq tV(s)$$

Submartingale Property: Step 3

- From: $sV(s) + sm_0(t-s) \geq tV(s)$
- Rearrange:

$$sm_0(t-s) \geq tV(s) - sV(s)$$

$$sm_0(t-s) \geq V(s)(t-s)$$

- Since $t > s$, we have $(t-s) > 0$, so divide by $(t-s)$:

$$sm_0 \geq V(s)$$

$$\boxed{m_0 \geq \frac{V(s)}{s}}$$

- This is exactly our assumption!

Submartingale Property: Conclusion

- Under the assumption $\frac{V(s)}{s} \leq m_0$ for all s:

$$E[Y_t|\mathcal{F}_s] = \frac{V(s) + m_0(t-s)}{t} \geq \frac{V(s)}{s} = Y_s$$

- Therefore, $Y_t = \frac{V(t)}{t}$ is a **submartingale**
- Interpretation:
  - The expected proportion of false discoveries *increases* over time
  - This holds as long as we haven't exceeded the m_0 threshold
  - BH procedure stops before violating this condition

Understanding Martingales: Fair Game Example

- Simple coin flip game:
  - Start with \$100
  - Win \$1 for Heads, Lose \$1 for Tails
  - Fair coin: $P(H) = P(T) = 1/2$
- Let X_n be fortune after n flips:

$$X_n = 100 + \sum_{i=1}^{n} Y_i \text{ where } Y_i = \pm 1$$

- Martingale property:

$$E[X_{n+1}|X_1, \ldots, X_n] = X_n$$

"Expected future fortune = current fortune"

Understanding Stopping Times: Examples

- Example 1: Stop when fortune reaches \$110

$$T_1 = \min\{n : X_n = 110\}$$

  - Is a stopping time
  - Only depends on past information

- Example 2: Stop at the last time fortune is \$90

$$T_2 = \max\{n : X_n = 90\}$$

  - Not a stopping time
  - Requires future information

Optional Stopping Theorem: Intuition via Fair Game

- Fair coin flipping game:
  - X_n: fortune after n flips (start at \$100)
  - Win/lose \$1 each flip
  - $X_n = 100 + \sum_{i=1}^{n} Y_i$, where $Y_i = \pm 1$
- Question: Can you win by clever stopping?
  - Stop when "ahead"?
  - Stop at "lucky" moments?

Example Stopping Rules

- Rule 1: Stop when reaching \$110

$$T_1 = \min\{n : X_n = 110\}$$

- Rule 2: Stop when "ahead" by \$10

$$T_2 = \min\{n : X_n - 100 = 10\}$$

- Rule 3: Stop after 100 flips

$$T_3 = 100$$

- All are valid stopping times!

Optional Stopping Theorem Statement

- For martingale $\{X_n\}$ and stopping time T:

$$E[X_T] = E[X_0]$$

- Under technical conditions:
  - T is finite almost surely
  - $E[|X_T|] < \infty$
  - $|X_n - X_{n-1}| \leq K$ (bounded increments)

OST in Action: Our Game

- For Rule 1 (stop at \$110):

$$E[X_{T_1}] = E[X_0] = 100$$

- Interpretation:
  - Can't expect to win
  - Even if stop at \$110 when possible
  - Balances with times you lose
- Key insight: No clever stopping rule can beat fairness!

Stopping Time

- BH threshold as stopping time:

$$T = \frac{qR(T)}{m}$$

(stops when threshold matches BH criterion)

- At stopping time:

$$Y_T = \frac{V(T)}{T} = \frac{V(T)m}{qR(T)}$$

- The assumption $\frac{V(s)}{s} \leq m_0$ is enforced by:
  - Stopping before Y_t would exceed m_0
  - Or restricting to paths where this holds

Applying Optional Stopping Theorem

- Since Y_t is a submartingale and T is a stopping time:

$$E[Y_T] \leq E[Y_1] = m_0$$

  for time 1.
- This gives us control over the expected FDP at the BH threshold
- Combined with $Y_T = \frac{V(T)m}{qR(T)}$, we can bound:

$$E\left[\frac{V(T)}{R(T)}\right] \leq q$$

(BH FDR control)

Reference

Storey, John D., Jonathan E. Taylor, and David Siegmund. "Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach." Journal of the Royal Statistical Society Series B: Statistical Methodology 66, no. 1 (2004): 187-205.