Sequential Multiple Hypothesis Testing with Type I Error Control

Alan Malek, Sumeet Katariya, Yinlam Chow, and Mohammad Ghavamzadeh

Hypothesis Testing, A/B Testing

1. Statisticaly rigorous to decide between
   (a) Null Hypothesis H_0 (e.g. $\mu_A = \mu_B$)
   (b) Alternative Hypothesis H_1 (e.g. $\mu_A \neq \mu_B$)
2. Hypothesis test: data $(x_1, \ldots, x_N) \mapsto$ {reject, fail to reject}
3. We require
   (a) Type I error (rejecting H_0 when it is true) $< \alpha$
   (b) Type II error (failing to reject H_0 when it is false) $< \beta$

Multiple Hypothesis Testing

1. Have m tests, $(H_{0,1}$ vs. $H_{1,1}), \ldots, (H_{0,m}$ vs. $H_{1,m})$
2. MHT procedure $\mathcal{P}$: $(p_1, \ldots, p_m) \mapsto \{\text{reject, fail to reject}\}^m$
3. Results summarized by

	not-rejected	rejected	total
H_0 true	U	V	m_0
H_0 false	W	S	$m - m_0$
total	$m - R$	R	m

4. Want U and S to be large, V and W to be small
5. No single notion of type I or II errors
6. Common goals

$$\text{FWER} := P(V \geq 1) \quad \text{and} \quad \text{FDR} := \mathbb{E}\left[\frac{V}{R \vee 1}\right]$$

7. Type II error versions

$$\text{FWER II} := P(W \geq 1) \quad \text{and} \quad \text{FNR} := \mathbb{E}\left[\frac{W}{(m-R) \vee 1}\right]$$

Error Notions

Definition 1 *$\mathcal{P}$ has an (f, q) error guarantee if*

$$\mathbb{E}\left[f(V, S)\right] \leq q \tag{1}$$

$\forall m, m_0$, distributions on S, and if true null p-values are marginally uniform.
If Eq. 1 holds only when $p^{\pi(1)}, \ldots, p^{\pi(m_0)}$ are i.i.d. uniform, we will say that $\mathcal{P}$ has an (f, q) error guarantee under independence (a weaker condition).

Common examples:

1. FWER: $f(V, S) = \mathbf{1}\{V > 0\}$
2. FDR: $f(V, S) = \frac{V}{(V+S) \vee 1}$

Common MHT procedures

Definition 2 (monotonic procedures) *$p^{(1)} \leq \ldots \leq p^{(m)}$ are in ascending order, $\alpha_1, \ldots, \alpha_m$ sequence of decision thresholds*

1. ***step-up*** *procedure rejects tests $(1), \ldots, (\max\{k : p^{(k)} \leq \alpha_k\})$*
2. ***step-down*** *procedure rejects tests $(1), \ldots, (\min\{k : p^{(k)} > \alpha_k\} - 1)$*

Most common examples:

Name	α	type	guarantee
Bonferroni	$\alpha_k = \frac{q}{m}$	both	FWER
Holm	$\alpha_k = \frac{q}{m-k}$	step-down	FWER
Hochberg	$\alpha_k = \frac{q}{m-k}$	step-up	FDR
Benjimini-Hochberg	$\alpha_k = \frac{k-q}{m}$	step-up	FDR

Sequential p-values

Definition 3 *A sequential p-value is a sequence of mappings $p_t : \mathcal{X}^t \to [0, 1]$ s.t.,* ***under the null hypothesis****,*

1. *(super-uniform) for any $\delta \in [0, 1]$ and any $t \geq 1$*

$$P\left(\sup_{s \leq t} p_s(X_1, \ldots, X_s) \leq \delta\right) \leq \delta \tag{2}$$

2. *(non-increasing) for any $\{x_t\}_{t \geq 1}$, for all t,*

$$p_t(x_1, \ldots, x_t) \geq p_{t+1}(x_1, \ldots, x_{t+1}). \tag{3}$$

Examples

1. SPRT: $\frac{1}{\sup_{n' \leq t} L_{n'}}$ where $L_n = \frac{\prod_{i=1}^n f_1(X_i)}{\prod_{i=1}^n f_0(X_i)}$ is the likelihood ratio
2. Test martingales: $\frac{1}{\sup_{n' \leq t} \Lambda_{n'}}$ where Λ_t is a positive supermartingale with $\mathbb{E}[\Lambda_0] = 1$.
   In general, for any such Λ_t, $P\left(\sup_t \boldsymbol{X}_t > b\right) \leq \frac{1}{b}$.

Sequential Conversion

Input: stopping time T, procedure $\mathcal{P}$, error q
Initialize $p_0^1, \ldots, p_0^m$ to 1 and $\mathcal{S}_0 = (0, \ldots, 0)$
For $t = 1, 2, \ldots,$

1. Set $\mathcal{S}_t = \mathcal{P}(p_{t-1}^1, \ldots, p_{t-1}^m)$
2. For each k, if $\mathcal{S}_t(k) = 0$, draw a sample and update p_t^k
3. Otherwise, set $p_t^k = p_{t-1}^k$
4. If T is reached or $\mathcal{S}_t = \{1, 1, \ldots, 1\}$, Break

Return decisions $\mathcal{S}_{T \vee t}$

Sequential Error Guarantee

Theorem 1 *Let $\mathcal{P}$ be a monotonic test procedure with a (f, q) guarantee. Then its sequential conversion $\mathcal{C}$ of $\mathcal{P}$ also has an (f, q) guarantee. That is,*

$$\mathbb{E}[f(V_T, S_T)] \leq q.$$

Furthermore, if $\mathcal{P}$ only has an independent (f, q) guarantee, then $\mathcal{C}$ only has an independent (f, q)-guarantee.

Sequential Calculator

1. Want a minimum sample size to guarantee power (β) and false discovery rate (α)
2. Using the Bonferroni correction,

$$\widetilde{N}^* = \max_{k \in \{1, \ldots, m\}} \inf\left\{t : \mathbb{P}(T^k \leq t | H_1^k) \geq 1 - \frac{\beta}{m}\right\}$$

where T^k is the stopping time of test k (with FWER guarantee α) that stops under the following condition:

$$\Lambda_t^k \geq m/\alpha, \ \forall k \in \{1, \ldots, m\}.$$

Experiments

1. 1000 Hypothesis tests $\mu = 0$ vs $\mu \neq 0$
2. Truth: 200 $\mu = 0$, 800 μ evenly spread $[-10, 10]$
3. Sequential conversion against Benjamini-Hochberg
4. Averaged over 1000 independent runs