**Ordered Hypothesis Testing Problem**

- Setup of Multiple Testing Problem: a sequence of hypotheses $H_1, \ldots, H_n$.
- $H_s = \{0, \ldots, s-1\}$ is true: $S = \{0, \ldots, s-1\}$, rejected $R = \{s \in \{0, \ldots, n\}: \mid S \cap H_s \mid \leq k \}$.
- FDP $= \frac{\mid R \mid}{t}$ be the False Discovery Proportion.
- FDR $= \frac{E[R]}{t}$ be the False Discovery Rate.
- A procedure that control FDR at level 0.1 produces a rejection set $S$ with roughly 90% being the true discoveries.

**Setup of Ordered Testing Problem**: $H_1, \ldots, H_n$ sorted via prior knowledge.

- Domain knowledge might be used to indicate which hypothesis is more "promising", i.e., less likely to be rejected.
- Heuristically, more focus should be put on "promising" hypotheses.

**A Unified Framework of Existing Procedures**

- Most existing Multiple Testing Procedures fall into the following framework:
  - Input: a sequence of p-values $p_1, \ldots, p_n$ associated with the hypotheses $H_1, \ldots, H_n$, usually assuming $p_i \sim U[0,1]$; for null hypothesis;$p_i \sim U[0,1]$; for null hypothesis;
  - Rejection Rule: the rejection set $S$ has the form $S = \{p_i \leq t \leq s\}$.
  - Choice of $s$ and $t$: maximize the number of rejection $\mid S \cap H_s \mid \leq k$.
  - With a target level $q$, where FDP $= \frac{\mid R \mid}{t}$ is a procedure-specific estimator of FDP.

**BH Procedure**: (Benjamini & Hochberg, 1997) $k \geq n$ and $\frac{\mid S \cap H_s \mid \leq k}{t}$.

**Storey’s BH Procedure** (Storey, et al., 2004) $k \geq n$ and $\frac{\mid S \cap H_s \mid \leq k}{t}$.

**Selective Sequept (SSS)** (Barber & Candes, 2015) is pre-fixed and $\frac{\mid S \cap H_s \mid \leq k}{t}$.

**Accumulation Test (AT)** (Li & Barber, 2015) $s = 1$ and for $h \geq 0$ with $\frac{\mid S \cap H_s \mid \leq k}{t}$.

**Seqstest** (Barber & Candes, 2015) AT with $h(x) = C(x > 1 - 1/\bar{C})$.

**Adaptive Seqstest and FDR Control**

- **Adaptive Seqstest (AS)**: $s$ is pre-fixed and $\frac{\mid S \cap H_s \mid \leq k}{t}$.
- Motivation: Similar to Storey’s correction of BH procedure. Notice that $\frac{\mid S \cap H_s \mid \leq k}{t}$ is the fraction of null hypotheses. Thus, $\frac{\mid S \cap H_s \mid \leq k}{t}$ is too conservative when $s$ is small. By contrast, $\frac{\mid S \cap H_s \mid \leq k}{t}$.

**FDR Control in Finite Samples**

**Theorem 1**. Assume that

- $\{p_i \leq \mu \}$ are independent of $\{p_i \leq \mu \}$;
- $\{p_i \leq \mu \}$ are i.i.d. with distribution function $F_i \equiv U[0,1]$.

Then AS controls FDR at level $q$.

**VCT Model and Asymptotic Power**

**Definition 1** (Varying Coefficient Two-groups (VCT) Model). An VCT($F_1, F_2$) model is a sequence of independent p-values $p_i \sim U[0,1]$ such that

- $p_i \sim \tau_i / \tau_{\mu}$ (for some distinct distributions $F_1$ and $F_2$),
- $p_i = f_i / \tau_{\mu}$ and $f_i$ are the null and non-null distributions and $\tau_{\mu}$ is the local non-null probability for $s \in [0,1]$.

For a VCT model, the Cumulative non-null fraction is defined as $F_{\mu} = \frac{1}{2}$.

**Theorem 2.** Consider a VCT model with $F_{\mu} = \frac{1}{2}$.

- $F_{\mu}$ is strictly decreasing and Lipschitz on $[0,1]$ with $F_{\mu} = \frac{1}{2}$.
- $f_i$ is the uniform distribution on $[0,1]$ and $f_i = F_{\mu}$ is strictly decreasing on $[0,1]$. Then $\frac{\mid S \cap H_s \mid \leq k}{t}$.

**Power Comparison: AS versus SS and AT**

**Real Data Example: GEOquery Data**

**Parameter Selection: $s$ and $k$**

- We take $s = q$ and $\lambda = 0.1$ as default and the left figure shows the simulated power in finite samples (with $\lambda = 0.1, \mu = 2, q = 0.01, \mid S \cap H_s \mid \leq 0.05$).
- $\lambda = 0.5$ is a rule of thumb and it is much more stable than a large $\lambda$, as suggested by theory.
- The choice of $s$ depends on the quality of ordering. Unless the ordering is very bad (either $\mid S \cap H_s \mid \leq 0.1$ or $\mid S \cap H_s \mid \leq 0.01$), $s = q$ gives a reasonable performance.

- We could try a grid of values for $s$, e.g., $(0.5, 0.25, 0.1)$, to improve the number of rejections. We will explore the validity of processes of this type in future researches.

**References**


