Tukey's Second-Level Significance Testing

Second-Level Significance Testing:

Higher criticism statistic:

$$HC_n^* = \max_{0<t\le \alpha_0} \frac{\hat{F}_n(t) - t}{\sqrt{t(1-t)}} \sqrt{n}$$

Sparse Mixtures.

Simple model with Equal Means:

$$H_0: X_i \overset{iid}{\sim} N(0,1) \quad \text{vs} \quad H_1: X_i \sim (1-\varepsilon) N(0,1) + \varepsilon N(\mu, 1)$$

The likelihood ratio: $L = \prod_{i=1}^{n} \left[(1-\varepsilon) + \varepsilon e^{\mu X_i - \frac{\mu^2}{2}} \right]$

ε controls the sparcity

μ controls the signal

$\varepsilon_n = n^{-\beta}$, $\quad \frac{1}{2} < \beta < 1$

$\mu_n = \sqrt{2r \log n}$, $\quad 0 < r < 1$

Let S be the number of non-nulls under H_1

$\Rightarrow ES = n\varepsilon$

a) $\beta = 1 \Rightarrow ES = 1 \Rightarrow$ "Needle in the haystack" problem

$\Rightarrow$ "Needle" should be longer than $\sqrt{2\log p}$

b) $\beta = \frac{1}{2} \Rightarrow ES = \sqrt{n} \Rightarrow$ "many small effects"

$\Rightarrow \theta = \frac{\|\mu\|^2}{\sqrt{2n}} \approx \frac{\mu^2}{\sqrt{2}} = \text{const}$

$\Rightarrow$ we investigate situation between a) and b)

Modification of Higher Criticism Test by Stepanova and Pavlenko (2014)

$$HC_{mod} = \max_{0<t<1} \sqrt{n} \frac{F_n(t) - t}{\sqrt{t(1-t)\, q(t)}},$$

where $q(t) = \log\log\left(\frac{1}{t(1-t)}\right)$

Multiple Testing Problem

We would like to make decision about each single hypothesis and we don't want to make many errors.

Example: n hypothesis : $1 \le i \le n$

$H_{0,i}$: gene i is null vs $H_{1,i}$: gene i is "non-null"

we are interested in each hypothesis

$$T_i = \frac{\text{ave}(\text{Cancer}) - \text{ave}(\text{Control})}{\text{Est St. Error}} \quad - \text{test function}$$

Under H_0: $T_i \sim t_{100}$.

p-value: $p_i = P(|t_{100}| > |T_i|)$

$p_1 \dots p_n$ - p-values corresponding to hypothesis $H_{01} \dots H_{0n}$.

$\Rightarrow$ we make decisions on each of hypothesis

Table representing multiple testing situation:

	accepted	rejected	total
TRUE	U	V	n_0
FALSE	T	S	$n - n_0$
total	$n-R$	R	n

$R = \#\{\text{rejected } H_{0,i}\}$ - is observed r.v. $\Leftrightarrow$

$\Leftrightarrow$ number of total discoveries (rejections of H_{0i})

$n - R = \#\{\text{accepted } H_{0,i}\}$

U, S - true decisions:

$U = \#\{H_{0i} \text{ is accepted} / H_{0i} \text{ is true}\}$

$S = \#\{H_{0i} \text{ is rejected} / H_{1i} \text{ is true}\}$

V, T - false desisions:

$V = \#\{H_{0i} \text{ is rejected} / H_{0i} \text{ is true}\} \Leftrightarrow$
$\Leftrightarrow$ false discoveries (Type I Error)
$T = \#\{H_{0i} \text{ is accepted} / H_{1i} \text{ is true}\} \Leftrightarrow$
$\Leftrightarrow$ false non-discoveries (Type II Error)
U, V, S, T are unobserved r.v.

How to define measure of the Type I Error in multiple testing?
There is no single way to extend the notion of Type I Error from single hypothesis testing to multiple case.
The classical way to define P(Type I Error) in multiple testing problem is definition of FWER - Familywise Error Rate:
$FWER = P(V \geq 1) \Leftrightarrow$ the probability that you made at least 1 false discovery.

FWER in a strong sense $\Leftrightarrow$ under all configurations of true and false hypotheses

We want to control FWER at low (α) level.

<u>Bonferroni's method</u>: We reject all $H_{0,i}$: $p_i \leq \frac{\alpha}{n}$.

<u>Only one assumption</u>: $p_i \sim U[0,1]$ under H_0.

<u>Theorem 1</u>: Bonferroni's method controls FWER at level α in the strong sense:
$$FWER \leq EV = \frac{n_0}{n}\alpha \qquad \left(\frac{n_0}{n} \leq 1\right)$$

⌐ Let

$$V_i = \begin{cases} 1, & \text{if } H_{0,i} \text{ is rejected} \\ 0, & \text{—"—} \end{cases}$$

$\mathcal{H}_0 = \{$ set of null hypotheses that are true $\} \Rightarrow$

$$\Rightarrow EV = \sum_{i \in \mathcal{H}_0} EV_i = \sum_{i \in \mathcal{H}_0} P(V_i = 1) = \sum_{i \in \mathcal{H}_0} \frac{\alpha}{n} = \frac{n_0}{n}\alpha$$

$$P(V \geq 1) \leq P(V \geq 1) + P(V \geq 2) + \ldots + P(V \geq n) = EV$$

Γ $V = \{0, 1, \ldots, n\}$ – number of FD (false discoveries)

$$EV = \sum_{i=0}^{n} i\, P(V=i) = \sum_{i=1}^{n} i\, P(V=i) =$$

$$= \sum_{i=1}^{n} P(V=i) + \sum_{i=2}^{n} P(V=i) + \ldots + P(V=n) =$$

$$= P(V \geq 1) + P(V \geq 2) + \ldots + P(V \geq n) \rfloor$$

Note: Bonf method might be conservative (It is redundant in the case when T_i are correlated)

Sidak's procedure   Assumption: independence of tests (p_i)

We reject all $H_{0,i}$: $p_i \leq \alpha_n$.

Then FWER =

$$= P(V \geq 1) = 1 - P(V = 0) = 1 - P\Big(\bigcap_{i \in \mathcal{H}_0} (V_i = 0)\Big) =$$

$$\overset{\text{inden.}}{=} 1 - \prod_{i=1}^{n_0} P(V_i = 0) = 1 - (1-\alpha_n)^{n_0} \overset{n_0 \leq n}{\leq} 1 - (1-\alpha_n)^n \leq \alpha$$

$$1 - (1-\alpha_n)^n \leq \alpha \Rightarrow \alpha_n \leq 1 - (1-\alpha)^{\frac{1}{n}}$$

$$\Rightarrow \alpha_n^S = 1 - (1-\alpha)^{\frac{1}{n}} \quad \text{compare with} \quad \alpha_n^{Bonf} = \frac{\alpha}{n}$$

$$\frac{\alpha_n^{Bonf}}{\alpha_n^S} = \frac{\frac{\alpha}{n}}{1 - (1-\alpha)^{\frac{1}{n}}} \approx \frac{\frac{\alpha}{n}}{1 - \left(1 - \frac{\alpha}{n} + \frac{\alpha^2}{2} \cdot \frac{1}{n}\left(\frac{1}{n} - 1\right)\right)} =$$

$$= \frac{\frac{\alpha}{n}}{\frac{\alpha}{n} - \frac{\alpha^2(1-n)}{2n^2}} = \frac{1}{1 - \frac{\alpha}{2} \cdot \underbrace{\frac{1-n}{n}}_{\approx -1}} \approx \frac{1}{1 + \frac{\alpha}{2}}$$

Weak Control

a testing procedure controls the FWER weakly if it is controls the FWER under the global null (i.e. when all $H_{0,i}$ are true)

Two-step procedure: (Fisher, 1934)

1. Global test for $H_0 = \bigcap_{i=1}^{n} H_{0,i}$
2. Test each hypothesis at level α

Example: 1. Reject H_0 if $\min p_i \le \frac{\alpha}{n}$

2. If H_0 is rejected $\Rightarrow$ reject $H_{0,i}$ if $p_i \le \alpha$

Note: This procedure does not control FWER in a strong sence.

Example: $X_i \sim N(\mu_i, 1)$ independent

$H_{0,i}: \mu_i = 0$

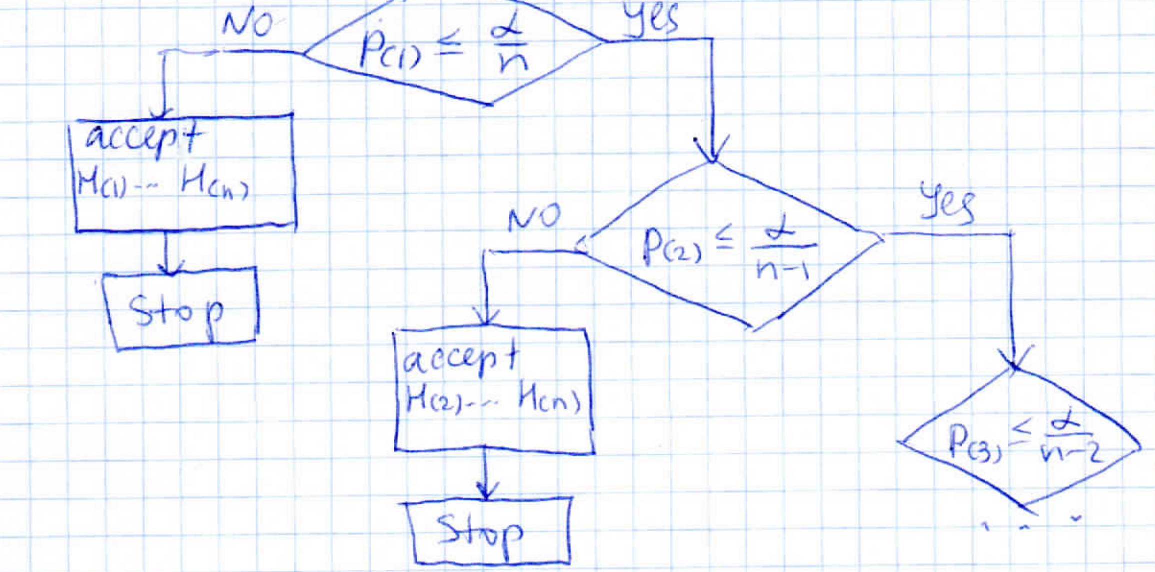
Let μ_1 is very large $\Rightarrow$ reject H_0

$\Rightarrow$ we apply α-level test to all the others

$\Rightarrow$ we maka $\approx \alpha n_0$ false discoveries :(

Holm's procedure (step-down procedure):

$$p_{(1)} \le p_{(2)} \le \dots \le p_{(n)}$$
$$H_{(1)} \quad H_{(2)} \quad \dots \quad H_{(n)}$$



Step i: if $p_{(i)} \le \frac{\alpha}{n-i+1}$ then reject $H_{(i)}$, go to the step $(i+1)$

else accept $H_{(i)} \dots H_{(n)}$ and stop

...

Step n: if $p_{(n)} \le \alpha$ then reject $H_{(n)}$

else accept $H_{(n)}$

The procedure stops the first time when

$$p_{(i)} \ge \alpha_i = \frac{\alpha}{n-i+1}$$

Note Holm's procedure is less conservative than Bonferroni's procedure (which rejects all $H_{0,i}$: $p_i < \frac{\alpha}{n}$)

Theorem 3. Holm's procedure controls the FWER strongly (without assumption of independence)

⌈ Let $i_0 = \min\{i : H_{0,i} \text{ is true}\}$

$$\underbrace{\begin{matrix} p_{(1)} \le \dots & \le p_{(i_0-1)} \le \\ H_{(1)} & H_{(i_0-1)} \end{matrix}}_{H_{1i} \text{ are true}} \underbrace{\begin{matrix} p_{(i_0)} \le & p_{(i_0+1)} \le & \dots & \le p_{(n)} \\ H_{(i_0)} & H_{(i_0+1)} & \dots & H_{(n)} \end{matrix}}_{H_{0i} \text{ are true and probably a few } H_{1i}}$$

We have n_0 hypotheses when H_{0i} are true.

$$\Rightarrow n - i_0 + 1 \ge n_0 \Rightarrow i_0 \le n - n_0 + 1$$

Then $$\{V \ge 1\} \subset \{p_{(i_0)} \le \underbrace{\frac{\alpha}{n-i_0+1}}_{\le \frac{\alpha}{n_0}}\} \subset \{p_{(i_0)} \le \frac{\alpha}{n_0}\}$$

$$\Rightarrow P\{V \ge 1\} \le P(\min_{i \in \mathcal{I}_0} p_i \le \frac{\alpha}{n_0}) \le \sum_{i \in \mathcal{I}_0} P(p_i \le \frac{\alpha}{n_0}) =$$

$$= \sum_{i \in \mathcal{I}_0} \frac{\alpha}{n_0} = n_0 \frac{\alpha}{n_0} = \alpha \quad ⌋$$