Lecture 31: False discovery rate

Outcomes for multiple hypothesis tests

Test m hypotheses, m_0 are truly null

	H_0 is true	H_0 is false	
Reject	✓		R
Fail to reject			
	m_0		m

$$FWER = P(V > 0)$$ (Probability of at least 1 type I error)

R = total # of rejections

$$FDP = \begin{cases} \frac{V}{R} & \text{if } R > 0 \\ 0 & \text{if } R = 0 \end{cases}$$

$$FDR = \mathbb{E}[FDP]$$

False discovery rate

Suppose we test m hypotheses, m_0 of which are truly null. Let V denote the number of type I errors, and R the total number of rejections.

$$FWER = P(V > 0) \quad FDR = \mathbb{E}[FDP]$$

① If $m_0 = m$, then $FWER = FDR$

Pf: Since $m_0 = m$, either $R = 0$, or $R > 0$ and $V = R$

$$\Rightarrow FDP = \begin{cases} 1 & R > 0 \\ 0 & R = 0 \end{cases}$$

$$\Rightarrow E[FDP] = P(R > 0) * 1 + P(R = 0) * 0$$
$$= P(R > 0) = P(V > 0) = FWER$$

② In general, $FDR \leq FWER$ (controlling FDR does not necessarily control FWER, but controlling FWER does control FDR)

Pf: $FDP \leq \mathbb{1}\{V > 0\}$

$$\Rightarrow FDR = \mathbb{E}[FDP] \leq \mathbb{E}[\mathbb{1}\{V > 0\}] = P(V > 0) = FWER$$

In R: p.adjust(...) methods include "BH", "BY", "Bonferroni", etc.

The Benjamini-Hochberg procedure

Suppose we test m null hypotheses $H_{0,1}, \ldots, H_{0,m}$. Let p_i be the corresponding p-value for test i.

- Order the p-values $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(m)}$
- Let $i^* = \max\left\{i : p_{(i)} < \frac{i\alpha}{m}\right\}$
- Reject $H_{0,(i)}$ for all $i \leq i^*$

i^* hypotheses rejected total

$\Rightarrow$ FDP $\approx \frac{m_0 i^* \alpha / m}{i^*}$

$\approx \frac{m_0}{m}\alpha \leq \alpha$

Claim: If the hypotheses are independent, BH controls FDR at level $\frac{m_0}{m}\alpha \leq \alpha$

Intuition: If H_0 is true, we expect p-values to be Uniform(0,1)

$i^* = \max\{i : p_{(i)} < \frac{i\alpha}{m}\} \Rightarrow$ reject p-values $p_i < \frac{i^*\alpha}{m}$

If H_0 is true, $P\left(p_i < \frac{i^*\alpha}{m}\right) = \frac{i^*\alpha}{m}$

we have m_0 the null hypothes $\Rightarrow$ expect # of the null hypothes that we reject is $\frac{m_0 i^* \alpha}{m}$

Summary

- BH controls FDR at level $\frac{m_0}{m}\alpha$
- If $m_0 = m$, then controlling FDR is equivalent to controlling FWER
- If $m_0 < m$, then controlling FDR provides more power to reject H_0 when H_0 is false