

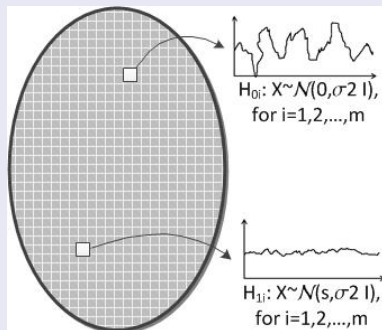
12. Multiple Hypothesis Testing

ECE 830, Spring 2014

Introduction

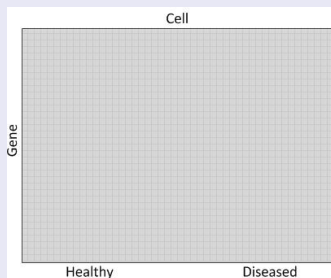
In many applications we consider multiple hypothesis test at the same time.

Example: FMRI (Functional Magnetic Resonance Imaging)



Brain image by FMRI with m voxels

Example: Microarrays



Is there a difference in the mean expression level in the healthy and diseased cells?

For the overall m genes that are expressed, we have for $i = 1, \dots, m$

$$H_{0i} : X_i \sim \mathcal{N}(0, 1)$$

$$H_{1i} : X_i \sim \mathcal{N}(\mu, 1), \quad \mu \neq 0$$

Suppose we have m tests and each is designed to guarantee $\mathbb{P}_{\text{FA}} \leq \alpha$. Then for any **one** test, the chance of a false alarm is α . However the probability of **at least one** false alarm among **all** the tests is much higher. **This is the multiple testing problem.**

For instance, imagine we perform a microarray study with $m = 10000$ wells, and choose $\mathbb{P}_{\text{FA}} = 5\%$. Then even if every single well corresponds to the null hypothesis (i.e. pure noise), then we will have about 500 wells that we falsely declare significant!

Consider m hypothesis tests (a family of tests): H_{0i} v.s. H_{1i} , where $i = 1, \dots, m$

Definition: FWER

The **family-wise error rate (FWER)** is the probability of one or more false alarms among any of the m tests:

$$\text{FWER} = \mathbb{P} \left(\bigcup_{i=1}^m \text{decide } H_{1i} \text{ when } H_{0i} \text{ is true} \right)$$

The Bonferroni Correction

In many applications we want to control the FWER. For example, we want to be confident that all detected voxels or genes are truly relevant.

Suppose we have two tests:

Test 1	Test 2
H_{01} v.s. H_{11}	H_{02} v.s. H_{12}
$t_1 \stackrel{H_{11}}{\geq} \gamma_1$ H_{01}	$t_2 \stackrel{H_{12}}{\geq} \gamma_2$ H_{02}
$\mathbb{P}(t_1 > \gamma_1 \mid H_{01}) \leq \alpha$	$\mathbb{P}(t_2 > \gamma_2 \mid H_{02}) \leq \alpha$

γ_1 and γ_2 are sets such that $\mathbb{P}_{\text{FA}} \leq \alpha$ in both cases.

Recall that for any two events A and B, we have

$$\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$$

with equality iff $A \cup B = \emptyset$. This is called the **union of events bound**, **union bound** or **Bonferroni's inequality**.

The Bonferroni Correction (cont.)

Thus,

$$\begin{aligned}\text{FWER} &= \mathbb{P}(t_1 > \gamma_1 \text{ or } t_2 > \gamma_2 \mid H_{01}, H_{02}) \\ &= \mathbb{P}(\{t_1 > \gamma_1 \mid H_{01}\} \cup \{t_2 > \gamma_2 \mid H_{02}\}) \\ &\leq 2\alpha\end{aligned}$$

More generally, if we have m tests and each has an individual $\mathbb{P}_{\text{FA}} \leq \alpha$, then $\text{FWER} \leq m\alpha$.

Bonferroni Correction

To guarantee a $\text{FWER} \leq \alpha$ for a family of m tests, we can set the $\mathbb{P}_{\text{FA}} \leq \frac{\alpha}{m}$ for each individual test.

Example: Threshold for multiple hypothesis testing

For $i = 1, \dots, m$,

$$H_{0i} : X_i \sim \mathcal{N}(0, 1)$$

$$H_{1i} : X_i \sim \mathcal{N}(\mu, 1), \quad \mu > 0$$

For the test $x_i \underset{H_{0i}}{\overset{H_{1i}}{\geq}} \gamma$, if we use

$$\gamma = Q^{-1}\left(\frac{\alpha}{m}\right)$$

instead of $\gamma = Q^{-1}(\alpha)$, then

$$\text{FWER} \leq \alpha$$

Example: Testing for differential expression in genes

Suppose we have 2638 genes or tests with $H_{0i} : X \sim \mathcal{N}(0, 1)$. If we want a FWER ≤ 0.05 , then the threshold

$$\gamma_{\text{FW}} = Q^{-1}\left(\frac{0.05}{2638}\right) = 4.119$$

will suffice.

For comparison, if we only consider one gene, then the threshold

$$\gamma_1 = Q^{-1}(0.05) = 1.6449$$

would guarantee $\mathbb{P}_{\text{FA}} \leq \alpha$.

Thus, the Bonferroni correction is a quite conservative threshold.

Multiple Testing When m Is Large

To get a better insight into this problem, let's consider the situation when m is very large and consider the composite tests for $i = 1, \dots, m$

$$H_{0i} : X_i \sim \mathcal{N}(0, 1)$$

$$H_{1i} : X_i \sim \mathcal{N}(\mu, 1), \quad \mu > 0$$

The Bonferroni Correction is conservative and the union bound may be too loose. **Can we do better?**

Consider the following bound on the tail of $\mathcal{N}(0, 1)$:

$$\frac{1}{\sqrt{2\pi}} \int_{\gamma}^{\infty} e^{-\frac{x^2}{2}} dx \leq \frac{1}{2} e^{-\frac{\gamma^2}{2}}$$

With this bound, we can conclude that the \mathbb{P}_{FA} of the test $x_i \underset{H_{0i}}{\overset{H_{1i}}{\gtrless}} \gamma$ satisfies $\mathbb{P}_{\text{FA}_i} \leq \frac{1}{2} e^{-\frac{\gamma^2}{2}}$.

Now the union bound implies

$$\text{FWER} \leq \sum_{i=1}^m \mathbb{P}_{\text{FA}_i} \leq \frac{m}{2} e^{-\frac{\gamma^2}{2}} = \frac{1}{2} e^{(-\frac{\gamma^2}{2} + \log m)}$$

To guarantee $\text{FWER} \leq \frac{1}{2}$, we must have

For large m , this is the Bonferroni threshold.

Can we control the FWER with a smaller threshold?

It can be shown that if $\omega_i \sim \mathcal{N}(0, 1)$, $i = 1, \dots, m$, then

$$\lim_{m \rightarrow \infty} \frac{\max_i \{\omega_i\}}{\sqrt{2 \log m}} \stackrel{a.s.}{=} 1.$$

Now note that

$$\text{FWER}(\gamma_m, m) =$$

=

So if we choose $\gamma_m < \sqrt{2 \log m}$, then

$$\lim_{m \rightarrow \infty} \text{FWER}(\gamma_m, m) = 1$$

Thus for large m , we must take $\gamma > \sqrt{2 \log m}$ if we want to keep $\text{FWER} < 1$

False Discover Rate (FDR) Control

FWER control guarantees that

$$\mathbb{P}(\text{one or more false alarms}) \leq \alpha$$

Perhaps the FWER constraint is too conservative. Alternatively, we could try to ensure that

$$\text{number of false alarms} \approx \alpha \cdot (\text{number of declared detections})$$

The declared detections includes both the false alarms and the correct detections. This constraint is less conservative than FWER. In other words, we could instead aim to guarantee that among all the tests for which we decide H_1 , only a small fraction α are false alarms.

Definition: false-discovery proportion

The **false-discovery proportion** is

$$\begin{aligned} \text{FDP} &= \frac{\text{number of false alarms}}{\text{number of discoveries}} \\ &= \frac{\text{number of false alarms}}{\text{number of false alarms} + \text{number of correct detections}} \end{aligned}$$

Definition: false-discovery rate

The **false-discovery rate** is

$$\text{FDR} = \mathbb{E}[\text{FDP}] = \mathbb{E} \left[\frac{\text{number of false alarms}}{\text{number of discoveries}} \right]$$

A **“discovery”** is made anytime we decide H_1 , whether it is the correct decision or not.

FDR Control

We can now aim to control the FDR :

$$\text{FDR} \leq \alpha$$

Generally, $\gamma_{\text{FDR}} \leq \gamma_{\text{FWER}}$ since it allows a small fraction of false-alarms.

So how to choose γ to guarantee $\text{FDR} \leq \alpha$?

Benjamini-Hochberg (BH) Threshold

The BH threshold is an adaptively selected threshold that controls the FDR $\leq \alpha$. It is generally lower than the Bonferroni threshold for FWER $\leq \alpha$.

The BH threshold is completed as follows. Assume

$$H_{0i} : X_i \sim \mathcal{N}(0, 1), \quad i = 1, \dots, m$$

Step 1. Compute the “p-values” for each observation: Let $X \sim \mathcal{N}(0, 1)$ and define

$$p_i := \mathbb{P}(X \geq x_i), \quad i = 1, \dots, m$$

Step 2. Sort p-values such that

$$p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(m)}$$

where $p_{(1)}$ represents the most extreme tail probability corresponding to the largest x_i and $p_{(m)}$ represents the least extreme tail probability corresponding to the smallest x_i .

Step 3. Set threshold according to

$$i^* := \max \left\{ i : p_i < \frac{i\alpha}{m} \right\} \text{ and } \gamma_{\text{BH}} = p_{i^*}$$

Then the test $x_i \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_{\text{BH}}$ has $\text{FDR} \leq \alpha$

Gaussian Tail Bound

The tail of the standard Gaussian $\mathcal{N}(0, 1)$ distribution satisfies the bound for any $t \geq 0$,

$$\frac{1}{\sqrt{2\pi}} \int_t^{\infty} e^{-\frac{x^2}{2}} dx \leq \min \left\{ \frac{1}{2} e^{-\frac{t^2}{2}}, \frac{1}{\sqrt{2\pi} t^2} e^{-\frac{t^2}{2}} \right\}$$

Proof: Consider

$$\begin{aligned} R &:= \frac{\frac{1}{\sqrt{2\pi}} \int_t^{\infty} e^{-\frac{x^2}{2}} dx}{e^{-\frac{t^2}{2}}} \\ &= \frac{1}{\sqrt{2\pi}} \int_t^{\infty} e^{-\frac{(x^2-t^2)}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_t^{\infty} e^{-\frac{(x-t)(x+t)}{2}} dx \end{aligned}$$

For the first bound, let $y = x - t$,

$$R = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{\frac{-y(y+2t)}{2}} dy \leq$$

For the second bound, note that

$$R \leq \frac{1}{\sqrt{2\pi}} \int_t^{\infty} e^{\frac{-2t(x-t)}{2}} dx$$

=

=

=