ECE 830 Fall 2011 Statistical Signal Processing

instructor: R. Nowak

Lecture 12: Multiple Hypothesis Testing

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

Example 1 *FMRI (Functional Magnetic Resonance Imaging)*

H_{0i}: X~$\mathcal{N}(0,\sigma2\ I)$,
for i=1,2,...,m

H_{1i}: X~$\mathcal{N}(s,\sigma2\ I)$,
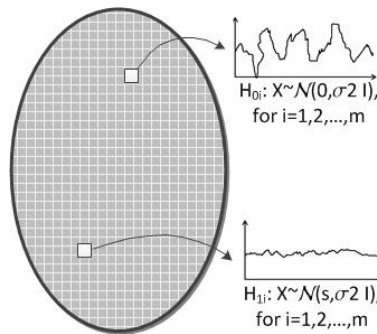for i=1,2,...,m

Figure 1: Brain image by FMRI with m voxels
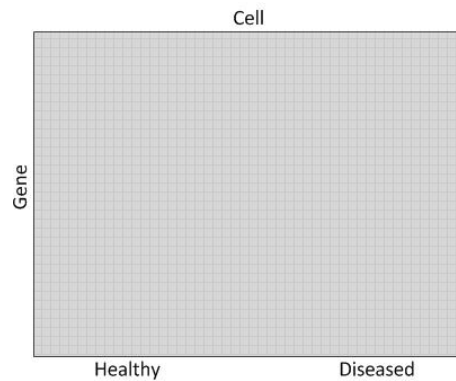
Example 2 *Microarrays*

Figure 2: Gene test plot

Is there a difference in the mean expression level in the healthy and diseased cells?
For the overall m genes that are expressed,

$$\mathbf{H}_{0i} : X_i \sim \mathcal{N}(0,1)$$
$$\mathbf{H}_{1i} : X_i \sim \mathcal{N}(\mu,1)$$
$$where\ i = 1, 2, ...m,\ \mu \neq 0$$

Suppose we have m tests and each is designed to guarantee $P_{FA} \leq \alpha$. Then for any one test, the chance of a false alarm is α. However the prob of <u>at least one</u> false alarm among <u>all</u> the tests is much higher. This is the multiple testing problem.

Consider m hypothesis tests (a family of tests): $\mathbf{H}_{0i}$ v.s. $\mathbf{H}_{1i}$, where $i = 1, 2, ..., m$

Definition 1 *The **family-wise error rate (FWER)** is the probability of one or more false alarms.*

$$FWER = \mathbb{P}(\bigcup_{i=1}^{m} \{\text{decide } \mathbf{H}_{1i} \text{ when } \mathbf{H}_{0i} \text{ is true}\})$$

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.

1   The Bonferroni Correction

Suppose we have two tests:

$$\mathbf{H}_{01} \text{ v.s. } \mathbf{H}_{11} \quad ; \quad \mathbf{H}_{02} \text{ v.s. } \mathbf{H}_{12}$$
$$t_1 \underset{H_{01}}{\overset{H_{11}}{\gtrless}} \gamma_1 \quad ; \quad t_1 \underset{H_{02}}{\overset{H_{12}}{\gtrless}} \gamma_1$$
$$\mathbb{P}(t_1 > \gamma_1 \mid \mathbf{H}_{01}) \leq \alpha \quad ; \quad \mathbb{P}(t_2 > \gamma_2 \mid \mathbf{H}_{02}) \leq \alpha$$

γ_1 and γ_2 are sets such that $P_{FA} \leq \alpha$ 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**.

Thus,

$$FWER = \mathbb{P}(t_1 > \gamma_1 \text{ or } t_2 > \gamma_2 \mid \mathbf{H}_{01}, \mathbf{H}_{02}) = \mathbb{P}(\{t_1 > \gamma_1 \mid \mathbf{H}_{01}\} \cup \{t_2 > \gamma_2 \mid \mathbf{H}_{02}\}) \leq 2\alpha$$

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

Definition 2 *Borferroni Correction*
To guarantee a $FWER \leq \alpha$ for a family of m tests, we can set the $\mathbb{P}_{FA} \leq \frac{\alpha}{m}$ for each individual test.

Example 3 *Threshold for multiple hypothesis testing*

$$\mathbf{H}_{0i} : X_i \sim \mathcal{N}(0, 1)$$
$$\mathbf{H}_{1i} : X_i \sim \mathcal{N}(\mu, 1)$$
$$\text{where } i = 1, 2, ...m, \ \mu \geq 1$$

For the test: $x_i \underset{H_{0i}}{\overset{H_{1i}}{\gtrless}} \gamma$, *if we use $\gamma = Q^{-1}(\frac{\alpha}{m})$ instead of $\gamma = Q^{-1}(\alpha)$, then $FWER \leq \alpha$*

Example 4 *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 $FWER \leq 0.05$, then the threshold $\gamma_{FW} = Q^{-1}(\frac{0.05}{2638}) = 4.119$ will suffice.
For comparison, if we only consider one gene, then the threshold $\gamma_1 = Q^{-1}(0.05) = 1.6449$ would guarantee $\mathbb{P}_{FA} \leq \alpha$.
Thus, the Bonferroni correction is a quite conservative threshold.

2   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:

$$\mathbf{H}_{0i} : X_i \sim \mathcal{N}(0, 1)$$
$$\mathbf{H}_{1i} : X_i \sim \mathcal{N}(\mu, 1)$$
$$\text{where } i = 1, 2, ...m, \ \mu \geq 0$$

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 P_{FA} of the test $x_i \overset{H_{1i}}{\underset{H_{0i}}{\gtrless}} \gamma$ satisfies $\mathbb{P}_{FA_i} \leq \frac{1}{2} e^{\frac{-\gamma^2}{2}}$. Now the union bound implies

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

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

$$\gamma \geq \sqrt{2 \log m}$$

For large m, this is the Bonferoni threshold.

Can we control the FWER with a smaller threshold? The answer, at least for large m, is no.
It can be shown that if $\omega_i \sim \mathcal{N}(0, 1)$, $i = 1, 2, ..., m$, then

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

$$\lim_{m \to \infty} FWER(\gamma_m, m) = 1 \text{ if } \gamma_m \leq \sqrt{2 \log m}$$

So, for large m, we must take $\gamma \geq \sqrt{2 \log m}$ if we want to keep $FWER \leq 1$

Example 5 *Consider the problem of "matched filtering." Suppose that we wish to detect the presence of a signal in noise, but the location/delay of the signal is unknown. More precisely, suppose the observation $x \in \mathbb{R}^n$ and consider the following hypothesis test*

$$H_0 : \quad x \quad \sim \quad \mathcal{N}(0, I)$$
$$H_1 : \quad x \quad \sim \quad \mathcal{N}(s_i, I), \quad \text{for one } i \in \{1, \ldots, m\}$$

The signals $\{s_i\}_{i=1}^{m}$ represent different shifted versions of the same basic waveform. This can be generalized to the case where the actual signal (if it is present) is close to, but not exactly equal to, one of the s_i.

3 False Discover Rate (FDR) Control

FWER control: $\mathbb{P}(\text{one or more false alarms}) \leq \alpha$. Perhaps the FWER constraint is too conservative. Alternative,

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

The declared detections includes both the false alarms and the correct detections. This constrain is less conservative than FWER. In other words, we could instead aim to guarantee that among all the tests for which we decide $\mathbf{H}_1$, only a small fraction α are false alarms.

Definition 3 *The **false-discovery proportion** is*

$$FDP = \frac{number\ of\ false\ alarms}{number\ of\ discoveries} = \frac{number\ of\ false\ alarms}{number\ of\ false\ alarms + number\ of\ correct\ detections}$$

Definition 4 *The **false-discovery rate** is*

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

A discovery is made anytime we decide $\mathbf{H}_1$, whether it is the correct decision or not. We can now aim to control the FDR as:

$$FDR \leq \alpha$$

Generally, $\gamma_{FDR} \leq \gamma_{FWER}$ since it allows a small fraction of false-alarms. Consequently, in general, FWER $\geq$ FDR. However, notice that if all the cases follow the null distribution, then there can be no correct detections (only false-positives), so in this case FDR=FWER.

4 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.

Step 1. Compute the "p-values" for each observation: Let t_i, $i = 1, \ldots, m$ denote the test statistics (e.g., log likelihood ratio statistics) and let T_0 denote a random variable whose distribution is that of the test statistic under the null distribution H_0. Define

$$p_i := \mathbb{P}(T_0 \geq t_i)\ ,\ i = 1, 2, ..., m$$

These are the so-called p-values. The size of the p-value for each t_i quantifies how typical or atypical it is under the null distribution. Smaller p-values correspond to test statistics that are improbable under the null.

Step 2. Sort p-values such that

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

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 $x_i \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_{\text{BH}}$ have $FDR \leq \alpha$

5 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

$$R \quad := \quad \frac{\frac{1}{\sqrt{2\pi}} \int\limits_{t}^{\infty} e^{\frac{-x^2}{2}} dx}{e^{-\frac{t^2}{2}}} \tag{1}$$

$$= \quad \frac{1}{\sqrt{2\pi}} \int\limits_{t}^{\infty} e^{\frac{-(x^2-t^2)}{2}} dx \tag{2}$$

$$= \quad \frac{1}{\sqrt{2\pi}} \int\limits_{t}^{\infty} e^{\frac{-(x-t)(x+t)}{2}} dx \tag{3}$$

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

$$R \quad = \quad \frac{1}{\sqrt{2\pi}} \int\limits_{0}^{\infty} e^{\frac{-y(y+2t)}{2}} dy \quad \leq \quad \frac{1}{\sqrt{2\pi}} \int\limits_{0}^{\infty} e^{\frac{-y^2}{2}} dy \quad = \quad \frac{1}{2} \tag{4}$$

For the second bound, note that

$$R \quad \leq \quad \frac{1}{\sqrt{2\pi}} \int\limits_{t}^{\infty} e^{\frac{-2t(x-t)}{2}} dx \tag{5}$$

$$= \quad \frac{1}{\sqrt{2\pi}} e^{t^2} \int\limits_{t}^{\infty} e^{-tx} dx \tag{6}$$

$$= \quad \frac{1}{\sqrt{2\pi}} e^{t^2} \frac{e^{-t^2}}{t} \tag{7}$$

$$= \quad \frac{1}{\sqrt{2\pi t^2}} \tag{8}$$