Large Scale Hypothesis Testing

2.1 Microarry Example

<u>Goal</u>: Identify "interesting genes"

Data:

$X = $ gene $\left\{ \begin{array}{c} \end{array} \right.$ X_{ij} — expression level
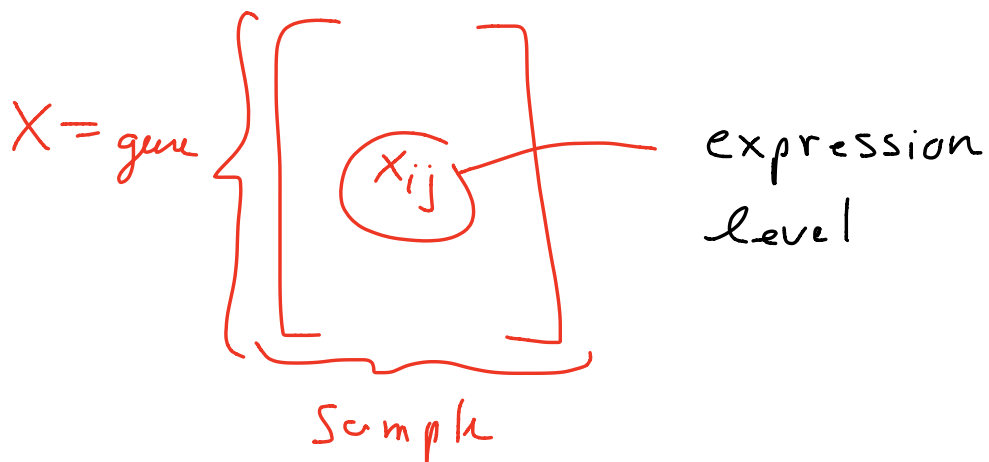
Sample

Partition the samples into controls versus infected spaces.

$$\overline{X}_i(k) = \text{ave. expression for}$$

$k = 1, 2$ controls / infected

The two sample t-stat for testing

gene i

$$t_i = \frac{\overline{X_i}(2) - \overline{X_i}(1)}{S_i}$$

$$S_i^2 = \frac{\underset{\text{Controls}}{\sum}(x_{ij} - \overline{x}(1))^2 + \underset{\text{infected}}{\sum}(x_{ij} - \overline{x}(2))^2}{100}\left(\frac{1}{50} + \frac{1}{52}\right)$$

Here we would test based on the T- distribution. But generally we want to consider everything in normal theory

$$z_i = \Phi^{-1}\left(F_{100}(t_i)\right)$$

In general we have $N \gg$ tests so we want to compare these $\{z_i, 1 \le i \le n\}$ to a general normal without making multiple testing mistakes.

The conservative approach: Bonfironni

Reject at α/N level.

More Modern approach: Bayesian Approach

2.2 Bayesian Approach

The Two group Model

- N cases either null or non-null with prob π_0, π_1

- z values having density either $f_0(z), f_1(z), F_0(Z), F_1(Z)$

- The mixture model discribing the model is

$$F(Z) = \pi_0 F_0(Z) + \pi_1 F_1(Z)$$

That is $z \sim F$

- Suppose we observe $z \in Z$ and wonder if it corresponds to the

null or non-null

$$\phi(Z) = \mathbb{P}(null \mid z \in Z)$$

$$= \pi_0 \frac{F_0(Z)}{F(Z)}$$

Bayes False Discovery rate

for Z

2.3 Empirical Bayesian Methods