1 Multiple Testing

<table>
<thead>
<tr>
<th>Decision</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Truth</td>
<td>$H_0$</td>
</tr>
<tr>
<td>$H_0$</td>
<td>$U$</td>
</tr>
<tr>
<td>$H_1$</td>
<td>$T$</td>
</tr>
<tr>
<td></td>
<td>$Q$</td>
</tr>
</tbody>
</table>

Table 1: $V$: False Positive (Type-I error) $T$: False Negative (Type-II error) $S$: True positive $U$: True Negative

Example: Let all hypothesis be null (i.e. $m_0 = m$) with $\alpha$ level 0.05.

\[
H_1, \ldots, H_m
\]
\[
Z_1, \ldots, Z_m \; (Z \in \{0, 1\})
\]
\[
Z = \sum_{i=1}^{m} Z_i
\]
\[
E(Z) = \sum_{i=1}^{m} E(Z_i)
\]
\[
= \sum_{i=1}^{m} \text{Pr} (\text{reject test } i \mid H_{0,i} \text{ true})
\]
\[
= m \alpha
\]

Therefore the number of false positive scales with the number of tests.

2 Controlling False Positives

1. Family-wise error rate [$\text{FWER}=P(V \geq 1)$]

2. False Discovery proportion [$\text{FDP} = \frac{V}{R \vee 1} ; R \vee 1 = \text{max}(R, 1)$]

3. Recall (sensitivity) = $\frac{S}{m_1}$

4. Specificity $\frac{U}{m_0}$

5. Precision = 1 - FDP
2.1 FWER control procedure

Give $0 \leq \alpha \leq 1$ and $\{p_1 \cdots p_m\}$, output a set of hypothesis $\{H_1 \cdots H_m\}$ such that $\text{FWER} \leq \alpha$

Bonferroni procedure: Reject hypothesis $i$ if $p_i \leq \frac{\alpha}{m}$

\[
\text{FWER} = P\left( \bigcup_{i=1}^{m} \text{reject } H_i \text{ at level } \frac{\alpha}{m} \right) \leq \sum_{i=1}^{m} P(\text{reject } H_i \text{ at level } \frac{\alpha}{m}) = \frac{m_0 \alpha}{m} \leq \alpha
\]

This will hold irrespective of null distribution or correlations among tests (strong FWER control).

2.2 False Discovery Rate

FDR = "Average proportion of false discoveries"

\[
\mathbb{E}(\text{FDP}) = \mathbb{E}\left( \frac{V}{R} \right) P(R > 0) \quad (11)
\]

Since $\text{FWER} \geq \text{FDR}$, any procedure that control FWER will control FDR. Define $\text{FDR}(t)$ as the FDR when we reject all null hypothesis where $p_i \leq t$.

\[
\text{FDR}(t) = \mathbb{E}\left( \frac{V(T)}{R(t) \lor 1} \right) \quad \sup_t \{t : \text{FDR}(t) \leq \alpha\}
\]

How do we set $\text{FDR}(t)$?

2.2.1 Benjamin-Hochberg FDR Control Procedure

Given $0 \leq \alpha \leq 1$ and ordered p-values resulting from your testing procedure $p_{(1)}, \cdots, p_{(m)}$.

\[
\hat{k} = \max_k \left\{ 1 \leq k \leq m \mid p_{(k)} \leq \frac{k}{m} \alpha \right\} \quad (12)
\]

Reject all hypothesis that correspond to $p_{(1)}, \cdots, p_{(\hat{k})}$. Otherwise reject none. Proof (non-rigorous):

\[
\text{FDR}(t_{BH}) = \mathbb{E}\left( \frac{\text{V}(t_{BH})}{\text{R}(t_{BH}) \lor 1} \right) \approx \frac{\mathbb{E}[\text{V}(t_{BH})]}{\mathbb{E}[\text{R}(t_{BH})]} \quad (14)
\]

\[
\text{R}(t_{BH}) = \mathbb{R} \quad (16)
\]

\[
\mathbb{E}(V \mid t_{BH}) \leq \frac{R}{m} \alpha m_0 \leq \frac{m}{m_0} \leq \alpha
\]

2
Therefore this procedure is more conservative than it needs to be. Under the null, $p \sim \text{Unif}(0,1)$.

\[
\text{FDR} \approx \frac{\mathbb{E}[V(t)]}{\mathbb{E}[R(t)]} \sqrt{1}
\]

\[
= \frac{m_0 t}{R(t)}
\]  

Only the denominator in eq (19) is observed. FDR control relies on independence or weak dependence among tests.

### 3 Linear Regression

Given: Output $y \in \mathbb{R}$ and Input $x \in \mathbb{R}^d$, goal is to find function $f : x \rightarrow y$ that predicts $y$ well. Assume

\[
f(x) = x'y = \sum_{i=1}^{d} x_i \beta_i
\]

is linear in the parameters $\beta$.

#### 3.1 Training with labeled data

Optimize a least squares criterion. Define residual sum of squares based on data labeled data $D$

\[
D = \{(x_i, y_i), i \in 1 \cdots n\}
\]

\[
RSS = \sum_{i=1}^{n} (y_i - x_i)^2
\]  

In vector/matrix notation eq 21 becomes

\[
RSS = \|Y - XB\|_2^2
\]  

Taking a model-based approach we have:

\[
y_i = \beta' x_i + \epsilon_i \quad \epsilon \sim \mathcal{N}(0, \sigma^2)
\]

\[
\ln \ell(\beta, \sigma^2) = -\frac{1}{2\sigma^2} \|Y - XB\|^2 - \frac{n}{2}\ln(\pi\sigma^2)
\]  

\[
\nabla_{\beta} \ln \ell(\hat{\beta}, \hat{\sigma}^2) = 0
\]

\[
\frac{\partial}{\partial \sigma^2} \ln \ell(\hat{\beta}, \hat{\sigma}^2) = 0
\]  

\[
(X'X)\beta = X'Y \quad \text{(Normal Equations)}
\]

\[
\hat{\beta} = (X'X)^{-1} X'Y
\]

\[
\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (y_i - x_i'\hat{\beta})^2}{n}
\]  

#### 3.2 Assumptions

1. $\mathbb{E}(\epsilon_i \mid X) = 0$

2. $\text{Cov}(\epsilon \mid X) = \sigma^2 I_n$ (i.e. homoskedastic error)
3.3 Propositions

1. $E(\hat{\beta} | X) = \beta$
2. $\text{Cov}(\hat{\beta}) = \sigma^2(X'X)^{-1}$
3. Gauss-Markov Thm. $\hat{\beta}$ has minimum variance among all competing linear unbiased estimators.

$$\hat{\beta}_{MLE} \sim N(\beta, \sigma^2(X'X)^{-1})$$

Prediction  
New data $y_* \implies y_* = \hat{\beta}'x_*$

Hypothesis Testing  $\beta_j = 0$

1. Wald test: $\hat{\beta}_j \sim N(\beta_j, \sigma^2(X'X)_j^{-1})$.
2. Likelihood ratio test: $2 \ln \ell(\hat{\beta}) - \ln \ell(\hat{\beta}_0)) \sim \chi^2_1$