# STA532 - hw10 sample solution

Sheng Jiang (sj156@duke.edu)

1. (a) $p_i$ admits the representation: $p_i = x_i z_i + u_i(1-z_i)$, where $x_i \overset{iid}{\sim} P_1$, $z_i \overset{iid}{\sim} Ber(\gamma)$, $u_i \overset{iid}{\sim} unif(0,1)$, and $x_i, z_i, u_i$ are independent.

By law of total probability,

$$\begin{aligned} Pr(p_1 < \tfrac{\alpha}{m}) &= Pr(p_1 < \tfrac{\alpha}{m} \mid z=1) Pr(z=1) + Pr(p_1 < \tfrac{\alpha}{m} \mid z=0) Pr(z=0) \\ &= Pr(x_1 < \tfrac{\alpha}{m}) \cdot \gamma + Pr(u_1 < \tfrac{\alpha}{m})(1-\gamma) \\ &= F_1(\tfrac{\alpha}{m})\gamma + \tfrac{\alpha}{m}(1-\gamma) \end{aligned}$$

(b). With Bonferroni method, $H_0$ is rejected if at least one $p_i < \frac{\alpha}{m}$.

$$\begin{aligned} Pr(\min_{1\le i\le m} p_i < \tfrac{\alpha}{m}) &= 1 - Pr(\min p_i \ge \tfrac{\alpha}{m}) = 1 - Pr(\bigcap_{i=1}^{m}\{p_i \ge \tfrac{\alpha}{m}\}) \\ &= 1 - Pr(p_1 \ge \tfrac{\alpha}{m})^m \quad \text{(by i.i.d.)} \\ &= 1 - [1 - F_1(\tfrac{\alpha}{m})\gamma - \tfrac{\alpha}{m}(1-\gamma)]^m \end{aligned}$$

(c). $[1 - F_1(\frac{\alpha}{m})\gamma - \frac{\alpha}{m}(1-\gamma)]^m = \exp\{m \log(1 - F_1(\frac{\alpha}{m})\cdot\gamma - \frac{\alpha}{m}(1-\gamma))\}$

For large $m$, $F_1(\frac{\alpha}{m}) \approx 0$, $\frac{\alpha}{m} \approx 0$, so $\log(1 - F_1(\frac{\alpha}{m})\gamma - \frac{\alpha}{m}(1-\gamma)) \approx - F_1(\frac{\alpha}{m})\gamma - \frac{\alpha}{m}(1-\gamma)$

$$Pr(\min_{1\le i\le m} p_i < \tfrac{\alpha}{m}) \approx 1 - \exp\{-F_1(\tfrac{\alpha}{m})\cdot m\gamma - \alpha(1-\gamma)\}$$

(d) Since $F_1(\cdot)$ is weakly increasing, prob of rejection increases as $\alpha$ increases.

Note $-F_1(\frac{\alpha}{m})\cdot m\gamma - \alpha(1-\gamma) = -m(F_1(\frac{\alpha}{m}) - \frac{\alpha}{m})\gamma - \alpha$,

therefore, when $F_1(\frac{\alpha}{m}) < \frac{\alpha}{m}$, prob of rejection increases as $\gamma$ decreases;

when $F_1(\frac{\alpha}{m}) = \frac{\alpha}{m}$, prob of rejection does not change with $\gamma$;

when $F_1(\frac{\alpha}{m}) > \frac{\alpha}{m}$, prob of rejection increases as $\gamma$ increases.

$\alpha$ is proportional to the decision boundary, when $\alpha$ becomes larger, it is more likely for $p_i$'s to be less than the threshold, i.e., one of the hypothesis is rejected.

$\gamma$ controls the likelihood of $p_i$ drawn from the alternative. If the alternative has more prob mass below the threshold than the null, that is, when $p_i$ is drawn from the alternative, $p_i$ is more likely ~~to~~ to be rejected than when it is drawn from the null.

(e). $F_1(\frac{\alpha}{m}) > \frac{\alpha}{m}$ and $m(F_1(\frac{\alpha}{m}) - \frac{\alpha}{m}) \to \infty$

2. • Under the null, $\sum_{i=1}^{m} Y_i^2 \sim \chi_m^2$, reject the null if $\sum_{i=1}^{m} Y_i^2 > Q(\chi_m^2, 1-\alpha)$ where $Q(\chi_m^2, 1-\alpha)$ denotes the $(1-\alpha)^{th}$ quantile of $\chi_m^2$.

• Under the null, p-value for observation $Y_i \sim N(0,1)$ is $p_i = 2(1-\Phi(|Y_i|))$, Using Bonferroni method, reject the null if any $p_i < \frac{\alpha}{m}$

• Fisher's procedure is to reject the null if $-2\sum_{i=1}^{m} \log p_i > Q(\chi_{2m}^2, 1-\alpha)$ where $Q(\chi_{2m}^2, 1-\alpha)$ denotes the $(1-\alpha)^{th}$ quantile of $\chi_{2m}^2$.

In (a), many minor effects.

In (b), needle in a haystack.

Simulation results are in the R script.

3. (a) $L(\{\theta_i\}) \propto e^{-\frac{1}{2}\sum_{i=1}^{m}(Y_i - \theta_i)^2}$

MLEs of $\{\theta_i\}$ are $\hat{\theta}_i = Y_i$

(b) Denote the values of hypothesis H as $\{\theta_{0,i}\}$.

Likelihood ratio between $\{\theta_i\}$ v.s. $\{\theta_{0,i}\}$ is

$$LR(\{\theta_i\}, \{\theta_{0,i}\}) = \frac{L(\{\theta_i\})}{L(\{\theta_{0,i}\})} = e^{-\frac{1}{2}\sum_{i=1}^{m}(\theta_i^2 - 2Y_i\theta_i)}$$

Then $-2\log LR(\{\theta_i\},\{\theta_{0,i}\}) = \sum_{i=1}^{m} \theta_i(\theta_i - 2Y_i)$.

Now replace $\{\theta_i\}$ with their MLE, $-2\log LR(\{\hat{\theta}_i\}, \{\theta_{0,i}\}) = -\sum_{i=1}^{m} Y_i^2$

Note, under $H$, $\sum_{i=1}^{m} Y_i^2 \sim \chi_m^2$ as $Y_i \overset{iid}{\sim} N(0,1)$.

Also note we reject $H$ for large values of $LR(\{\hat{\theta}_i\}, \{\theta_{0,i}\})$.

Then, we reject $H$ if $\sum_{i=1}^{m} Y_i > Q(\chi_m^2, 1-\alpha)$ where $Q(\chi_m^2, 1-\alpha)$ denotes $(1-\alpha)^{th}$ quantile of $\chi_m^2$. This procedure is level-$\alpha$.

(c). (i). For some $u \in (0,1)$, $Pr(p_j < u) = Pr(2\Phi(-|Y_j|) < u)$

$$= Pr(|Y_j| > \Phi^{-1}(1-\tfrac{u}{2}))$$

$$= 2Pr(Y_j > \Phi^{-1}(1-\tfrac{u}{2})) \qquad \text{note } Y_j \sim N(0,1)$$

$$= u.$$

(ii) Let $E_j = \{H_j \text{ is rejected at level } \alpha/m\}$, then $Pr(E_j \mid H_j) = \frac{\alpha}{m}$.

$$Pr(\bigcup_{j=1}^{m} E_j \mid H_0) \le \sum_{j=1}^{m} Pr(E_j \mid H_0) \qquad \text{union bound, Boole's inequality}$$

$$= \sum_{j=1}^{m} \frac{\alpha}{m} = \alpha.$$

4. (a). With the knowledge of $\gamma$ and $b$, we can compute the marginal CDF of p-vals

$$Pr(p_i < t) = Pr(p_i < t \mid H_0)\, Pr(H_0) + Pr(p_i < t \mid H_1)\, Pr(H_1)$$

$$= t\cdot(1-\gamma) + \left(b\int_0^t (1-x)^{b-1}\, dx\right)\cdot\gamma$$

$$= t\cdot(1-\gamma) + (1-(1-t)^b)\gamma$$

Then BH FDR control finds maximum $t$ s.th. $\frac{t}{t(1-\gamma)+(1-(1-t)^b)\gamma} \le \alpha$.

As $\gamma$ is known, null proportion $1-\gamma$ is known, and we can exactly control FDR by choosing the maximum $t_\alpha$ s.th. $\frac{t_\alpha(1-\gamma)}{t_\alpha(1-\gamma)+(1-(1-t_\alpha)^b)\gamma} \le \alpha$.

The modified B-H procedure is to reject $H_i$'s whose p-val $p_i < t_\alpha$.

Note when $b=1$, $P_0 = P_1$ and $\gamma = 0$, all discoveries are "false", it's impossible to do FDR control at level $\alpha < 1$.

(b). Let $H \sim Ber(\gamma)$. If $H=1$, $p_i \sim P_1$; if $H=0$, $p_i \sim P_0$.

By law of total expectation:

$$\mathbb{E}p_1 = \mathbb{E}[\mathbb{E}(p_1 \mid H)] = \mathbb{E}(p_1 \mid H=0)\cdot Pr(H=0) + \mathbb{E}(p_1 \mid H=1)\cdot Pr(H=1) = (1-\gamma)\tfrac{1}{2} + \gamma \cdot \tfrac{1}{b+1}$$

By law of total variance: $V(p_1) = \mathbb{E}[V(p_1 \mid H)] + V[\mathbb{E}(p_1 \mid H)]$.

$V(p_1 \mid H=0) = \frac{1}{12}$, as $p_1 \mid H=0 \sim unif[0,1]$;

$V(p_1 \mid H=1) = \frac{b}{(b+1)^2(b+2)}$, as $p_1 \mid H=1 \sim Beta(1,b)$

Then $\mathbb{E}[V(p_1 \mid H)] = V(p_1 \mid H=0)\,Pr(H=0) + V(p_1 \mid H=1)\,Pr(H=1) = \frac{1-\gamma}{12} + \frac{b\gamma}{(b+1)^2(b+2)}$

$\mathbb{E}(p_1 \mid H=0) = \frac{1}{2}$, $\mathbb{E}(p_1 \mid H=1) = \frac{1}{b+1}$

$\mathbb{E}[(\mathbb{E}[p_1 \mid H])^2] = [\mathbb{E}(p_1 \mid H=0)]^2 \cdot Pr(H=0) + (\mathbb{E}(p_1 \mid H=1))^2\, Pr(H=1) = \frac{1}{4}(1-\gamma) + \frac{1}{(b+1)^2}\gamma$

$$V(\mathbb{E}(p_1 \mid H)) = \mathbb{E}[(\mathbb{E}(p_1 \mid H))^2] - (\mathbb{E}[\mathbb{E}(p_1 \mid H)])^2 = \tfrac{1}{4}(1-\gamma) + \tfrac{1}{(b+1)^2}\gamma - \left((1-\gamma)\tfrac{1}{2} + \gamma\cdot\tfrac{1}{b+1}\right)^2$$
$$= \gamma(1-\gamma)\left(\tfrac{1}{4} - \tfrac{b}{(b+1)^2}\right)$$

To sum up, $V(p_1) = \frac{1-\gamma}{12} + \frac{b\gamma}{(b+1)^2(b+2)} + \gamma(1-\gamma)\left(\frac{1}{4} - \frac{b}{(b+1)^2}\right)$

By method of moment, estimator $(\hat{\gamma}, \hat{b})$ for $(\gamma, b)$ solves

$$\begin{cases} (1-\gamma)\frac{1}{2} + \gamma\cdot\frac{1}{b+1} = \frac{1}{m}\sum_{i=1}^{m} p_i \\ \frac{1-\gamma}{12} + \frac{b\gamma}{(b+1)^2(b+2)} + \gamma(1-\gamma)\left(\frac{1}{4} - \frac{b}{(b+1)^2}\right) = \frac{1}{m}\sum_{i=1}^{m}(p_i - \bar{p}_m)^2 \end{cases}$$

Therefore, when $\gamma$ and $b$ are unknown, choose the maximum $t$ s.th. $\frac{t(1-\hat{\gamma})}{t(1-\hat{\gamma}) + (1-(1-t)^{\hat{b}})\hat{\gamma}} \le \alpha$. Denote the solution as $t^*$.

The modified B-H procedure is to reject $H_i$'s whose p-val $p_i < t^*$.

(c) See the attached R script.