STA 532 Homework10

Jae Hyun Lee, jl914

10 April, 2020

HW10 for STA-532

1.

(a)

$$
\begin{aligned}
&H \sim binary(\gamma) \\
&p_1 \mid H = 0 \sim unif(0,1) \to Pr(p_1 < \frac{\alpha}{m} \mid H = 0) = \frac{\alpha}{m} \\
&p_1 \mid H = 1 \sim P_1 \to Pr(p_1 < \frac{\alpha}{m} \mid H = 1) = F_1(\frac{\alpha}{m}) \\
&Pr(p_1 < \frac{\alpha}{m}) = Pr(p_1 < \frac{\alpha}{m} \mid H = 0)Pr(H = 0) + Pr(p_1 < \frac{\alpha}{m} \mid H = 1)Pr(H = 1) \\
&= (1-\gamma)\frac{\alpha}{m} + rF_1(\frac{\alpha}{m})
\end{aligned}
$$

(b)

Under H_0,

$$
\begin{aligned}
p_1 \cdots p_m &\sim^{iid} unif(0,1) \\
&Let\ p_{(1)}\ be\ the\ smallest\ pvalue \\
Pr(p_{(1)} < \frac{\alpha}{m}) &= 1 - Pr(p_{(1)} > \frac{\alpha}{m}) \\
&= 1 - \prod Pr(p_{(i)} > \frac{\alpha}{m}) \\
&= 1 - (1 - \frac{\alpha}{m})^m
\end{aligned}
$$

(c)

$$
\begin{aligned}
log(1 - \frac{\alpha}{m}) \sim -\frac{\alpha}{m} &\to (1 - \frac{\alpha}{m})^m \sim e^{-\alpha} \\
&\to 1 - (1 - \frac{\alpha}{m})^m \sim 1 - e^{\alpha}
\end{aligned}
$$

(d)

At (a), we find $Pr(p_1 < \frac{\alpha}{m}) = 1 - Pr(p_1 > \frac{\alpha}{m})$. At random α, γ, as we have shown in (b), $p_{(1)}$, the smallest p-value, reject H with probability $1 - \prod Pr(p_i > \frac{\alpha}{m}) = 1 - (1 - (1-\gamma)\frac{\alpha}{m} - \gamma F_1(\frac{\alpha}{m}))^m$.

For α, since F_1 is non-decreasing function, $1 - (1 - (1-\gamma)\frac{\alpha}{m} - \gamma F_1(\frac{\alpha}{m}))^m$ decrease as α increase. In addition, for fixed α, the relationship between γ and rejection probability depends on magnitude of $\frac{\alpha}{m}$ and $F_1(\frac{\alpha}{m})$. If $\frac{\alpha}{m} > F_1(\frac{\alpha}{m})$, then as γ decrease, the rejection probability increase.

(e)

The condition is that $F_1(\frac{\alpha}{m}) > \frac{\alpha}{m}$ for every m. Then,

$$
\begin{aligned}
&1-(1-\gamma)\frac{\alpha}{m}-\gamma F_1(\frac{\alpha}{m}) < 1-\frac{\alpha}{m} \\
&\rightarrow (1-(1-\gamma)\frac{\alpha}{m}-\gamma F_1(\frac{\alpha}{m}))^m < e^{-\alpha} \\
&\rightarrow 1-(1-(1-\gamma)\frac{\alpha}{m}-\gamma F_1(\frac{\alpha}{m}))^m > 1-e^{-\alpha} \\
&\rightarrow which\ indicates\ increase\ of\ rejection\ probability\ as\ m\ increase
\end{aligned}
$$

2.

Under H,

$$
\begin{aligned}
&Y_j-\theta_j/\sigma = Z_j = Y_j \sim N(0,1) \\
&\rightarrow Y_j^2 \sim \chi_1^2 \\
&\sum_{j=1}^{m} Y_j^2 \sim \chi_m^2
\end{aligned}
$$

by property of chi-sqaure distribution.

(a)

```
set.seed(532)
m <- 100
K <- c(1,4,16,64)

sd <- matrix(rep(K,m),ncol =4,byrow = T)^(1/2)
theta <- matrix(rnorm(m*4, mean = 0, sd = 0.1),ncol = 4)*sd
cv <- qchisq(p = 0.95, df = 100,lower.tail = T)
cv2 <- qchisq(p = 0.95, df = 200,lower.tail = T)

chi_test <- rep(NA,4)
Bonf_test <- rep(NA,4)
Fish_test <- rep(NA,4)

for(i in 1:4){
  Y <- abs(MASS::mvrnorm(n = 1000, mu = theta[,i], Sigma = diag(rep(1,m),nrow = m)))
  chi_test[i] <- mean(apply(Y,1,function(x){sum(x^2)})>cv)
  Bonf_test[i] <- mean(apply(Y,1,function(x){sum(2*pnorm(x,lower.tail = F)<0.0005)})>0)
  Fish_test[i] <- mean(apply(Y,1,function(x){sum(-2*log(2*pnorm(abs(x),lower.tail = F)))})>cv2)
}
kable(rbind(chi_test,Bonf_test,Fish_test),col.names = paste("K = ",K),caption = "Probability of rejecti
```

Table 1: Probability of rejecting null

	K = 1	K = 4	K = 16	K = 64
chi_test	0.056	0.074	0.323	0.994
Bonf_test	0.071	0.064	0.123	0.550
Fish_test	0.062	0.078	0.311	0.991

(b)

```
K <- c(1,3,5,7)
chi_test2 <- rep(NA,4)
Bonf_test2 <- rep(NA,4)
Fish_test2 <- rep(NA,4)
for(i in 1:4){
  theta <- c(K[i],rep(0,m-1))
  Y <- MASS::mvrnorm(n = 1000, mu = theta, Sigma = diag(rep(1,m),nrow = m))
  chi_test2[i] <- mean(apply(Y,1,function(x){sum(x^2)})>cv)
  Bonf_test2[i] <- mean(apply(Y,1,function(x){sum(2*pnorm(abs(x),sd = 1, lower.tail = F)<0.0005)})>0)
  Fish_test2[i] <- mean(apply(Y,1,function(x){sum(-2*log(2*pnorm(abs(x),sd = 1, lower.tail = F)))})>cv2
}
kable(rbind(chi_test2,Bonf_test2,Fish_test2),col.names = paste("K = ",K),caption = "Probability of reje
```

Table 2: Probability of rejecting null

	K = 1	K = 3	K = 5	K = 7
chi_test2	0.057	0.161	0.499	0.887
Bonf_test2	0.053	0.355	0.949	1.000
Fish_test2	0.057	0.144	0.395	0.752

3.

(a)

$$\begin{aligned}
\tilde{y} &\sim N(\tilde{\theta}, I) \\
\rightarrow P(\tilde{y} \mid \tilde{\theta}) &= (2\pi)^{-\frac{m}{2}} exp\{-\frac{1}{2}(\tilde{y}-\tilde{\theta})^T(\tilde{y}-\tilde{\theta})\} \\
&= (2\pi)^{-\frac{m}{2}} exp\{-\frac{1}{2}(\tilde{\theta}^T\theta - 2\tilde{\theta}\tilde{y} + \tilde{y}^T\tilde{y})\} \\
l(\tilde{\theta}, \tilde{y}) &= c - \frac{1}{2}(\tilde{\theta}^T\theta - 2\tilde{\theta}\tilde{y} + \tilde{y}^T\tilde{y}) \\
\frac{d}{d\tilde{\theta}} l(\tilde{\theta}, \tilde{y}) &= -2\tilde{\theta} + 2\tilde{y} = 0 \\
\rightarrow \tilde{\theta}_{mle} &= \tilde{y}
\end{aligned}$$

(b)

-2log likelihood ratio statistics

$$-2log(\frac{L(\theta_0, \tilde{y})}{L(\theta_{mle}, \tilde{y})}) = -2log(\frac{(2\pi)^{-\frac{m}{2}} exp\{-\frac{1}{2}(\tilde{y}-\theta_0)^T(\tilde{y}-\theta_0)\}}{(2\pi)^{-\frac{m}{2}} exp\{-\frac{1}{2}(\tilde{y}-\theta_{mle})^T(\tilde{y}-\theta_{mle})\}}) = -2log(exp\{-\frac{1}{2}\tilde{y}^T\tilde{y}\}) = \tilde{y}^T\tilde{y} = \sum y_i^2$$

If $\sum y_i^2 > c$ reject H_0, O.W do not reject H_0. Since, $\sum y_i^2 \sim \chi^2_m, c = \chi^2_{m,(1-\alpha)}$.

(c)

i) If $\theta_j = 0, Y_j \sim N(0,1)$,

$$\begin{aligned}
Pr(- \mid Y_j \mid \leq \alpha) &= \Phi(- \mid Y_j \mid) = \alpha \\
Pr(p_j \leq \alpha) &= Pr(2 \times \Phi(- \mid Y_j \mid) \leq \alpha) \\
&= Pr(\Phi(- \mid Y_j \mid) \leq \frac{\alpha}{2}) \\
&= Pr(- \mid Y_j \mid \leq \Phi^{-1}(\frac{\alpha}{2})) \\
&= Pr(- \mid Y_j \mid \leq Z_{\frac{\alpha}{2}})
\end{aligned}$$

Since, $Y_j \sim N(0,1)$, $Pr(- \mid Y_j \mid \leq Z_{\frac{\alpha}{2}}) = \alpha$. Thus $Pr(p_j \leq \alpha) = \alpha$ which indicates that p_j has uniform distribution.

ii) reject H_0 if any H_j are rejected at level $\frac{\alpha}{m}$.

$$\begin{aligned}
Pr(rej\ H_0 \mid H_0) &= Pr((P_1 < \frac{\alpha}{m})\ or\ (P_2 < \frac{\alpha}{m})\ \cdots (P_m < \frac{\alpha}{m})) \\
&= Pr(\cup\{P_i < \frac{\alpha}{m} \mid H_0\}) \\
&\leq \sum Pr(\{P_i < \frac{\alpha}{m} \mid H_0\})\ because\ of\ intersection\ term \\
&= m\frac{\alpha}{m} = \alpha\ because\ P_i \sim N(0,1)
\end{aligned}$$

4.

(a)

$$\begin{aligned}
&H \sim binary(\gamma) \\
&p \mid H = 0 \sim unif(0,1) \\
&p \mid H = 1 \sim beta(1,b)
\end{aligned}$$

$$\begin{aligned}
FDP &= \frac{\sum Pr(p_i < \alpha_E\ and\ H_i = 0)}{\sum Pr(p_i < \alpha_E)} \\
&= \frac{\sum Pr(p_i < \alpha_E \mid H_i = 0) Pr(H_i = 0)}{\sum Pr(p_i < \alpha_E)} \\
&= \frac{\sum (1-\gamma)\alpha_E}{\sum (1-\gamma)\alpha_E + \gamma F_i(\alpha_E)}
\end{aligned}$$

where

$$\begin{aligned}
F_1(\alpha_E) &= \int_0^{\alpha_E} \frac{\Gamma(b+1)}{\Gamma(b)\Gamma(1)} \times x^{1-1}(1-x)^{b-1} dx \\
&= \frac{\Gamma(b+1)}{\Gamma(b)\Gamma(1)} \times \int_0^{\alpha_E} (1-x)^{b-1} dx \\
&= \frac{\Gamma(b+1)}{\Gamma(b)\Gamma(1)} \frac{1}{b}(1-(1-\alpha_E)^b) = 1-(1-\alpha_E)^b
\end{aligned}$$

So,

$$
\begin{aligned}
FDR &= \frac{m(1-\gamma)\alpha_E}{m[(1-\gamma)\alpha_E + \gamma(1-(1-\alpha_E)^b)]} \\
&= \frac{(1-\gamma)\alpha_E}{(1-\gamma)\alpha_E + \gamma(1-(1-\alpha_E)^b)}
\end{aligned}
$$

To control FDR, we need to choose α_E which satisfy $\frac{(1-\gamma)\alpha_E}{(1-\gamma)\alpha_E+\gamma(1-(1-\alpha_E)^b)} < \alpha$ because we know γ, b, we don't need to use approximate bound. So modified BH method is $H_{0,j} : \theta_j = 0$ if $p_j < \alpha_E$ where α_E is the max value for which $\frac{(1-\gamma)\alpha_E}{(1-\gamma)\alpha_E+\gamma(1-(1-\alpha_E)^b)} < \alpha$. After, sort p_j as previous way, $p_{(k)} < \alpha_E$ if $\frac{(1-\gamma)\alpha_E}{(1-\gamma)\alpha_E+\gamma(1-(1-\alpha_E)^b)} < \alpha$

(b)

$$
\begin{aligned}
&H \sim binary(\gamma) \\
&p \mid H=0 \sim unif(0,1) \rightarrow E(p \mid H=0) = \frac{1}{2}, E(p^2 \mid H=0) \\
&p \mid H=1 \sim beta(1,b) \rightarrow E(p \mid H=1) = \frac{1}{(b+1)}, E(p^2 \mid H=1) = \frac{2}{(b+2)(b+1)}
\end{aligned}
$$

$$
\begin{aligned}
E(p_1) &= E(E(p_1 \mid H)) \\
&= E(p_1 \mid H=0)Pr(H=0) + E(p_1 \mid H=1)Pr(H=1) \\
&= (1-\gamma)E(p_1 \mid H=0) + \gamma E(p_1 \mid H=1) \\
&= \frac{1}{2}(1-\gamma) + \frac{\gamma}{b+1} = \gamma(\frac{1}{b+1} - \frac{1}{2}) + \frac{1}{2}
\end{aligned}
$$

$$
\begin{aligned}
E(p_1^2) &= E(E(p_1^2 \mid H)) \\
&= (1-\gamma)E(p_1^2 \mid H=0) + \gamma E(p_1^2 \mid H=1) \\
&= (1-\gamma)\frac{1}{3} + \frac{2\gamma}{(b+2)(b+1)} = \gamma(\frac{2}{(b+1)(b+2)} - \frac{1}{3}) + \frac{1}{3}
\end{aligned}
$$

$$
\begin{aligned}
V(p_1) &= E(p_1^2) - E(p_1)^2 \\
&= (1-\gamma)\frac{1}{3} + \frac{2\gamma}{(b+2)(b+1)} - \frac{1}{4}(1-\gamma)^2 - \frac{\gamma(1-\gamma)}{(b+1)} - \frac{\gamma^2}{(b+1)^2}
\end{aligned}
$$

$$
E(p_1^k) = (1-\gamma)\frac{1}{k+1} + \gamma\frac{\Gamma(k+1)}{\Gamma(b+k+1)} = \gamma(\frac{\Gamma(k+1)}{\Gamma(b+k+1)} - \frac{1}{k+1}) + \frac{1}{k+1}
$$

We can solve k equations

$$
\begin{bmatrix} \frac{1}{n}\sum p_i \\ \frac{1}{n}\sum p_i^2 \\ \cdots \\ \frac{1}{n}\sum p_i^k \end{bmatrix} = \begin{bmatrix} equation1 \\ equation2 \\ \cdots \\ equation3 \end{bmatrix} \quad regarding\ \gamma, b
$$

and then, get estimation of solution $\hat{\gamma}, \hat{b}$. Finally we can use this estimation $\hat{\gamma}, \hat{b}$ instead of γ, b and plug-in them into $\frac{(1-\hat{\gamma})p_{(k)}}{(1-\hat{\gamma})p_{(k)}+\hat{\gamma}(1-(1-p_{(k)})^{\hat{b}})} < \alpha$.