STA 532 HW10 Due Apr 20 Yi Mi ym154

1. (a) $Pr(p_1 < \alpha/m) = \int_0^{\alpha/m} (1-y)p_0 + y p_1 \, dt$

$= (1-y)\frac{\alpha}{m} + y F_1(\alpha/m)$

(b) $H_0 : y=0$

$Pr(\text{rej } H_0 | H_0) = Pr(\text{rej any } H_0 | H_0)$

$= Pr(\text{any } p_j < \alpha/m | H_0)$

$= 1 - Pr(\text{all } p_j > \alpha/m | H_0)$

$= 1 - \prod Pr(p_j > \alpha/m)$

(c) $\log(1 - Pr(\text{rej } H_0 | H_0)) = m \log(1 - (1-y)\alpha/m - y F_1(\alpha/m))$

For small x, $\log(1-x) \approx -x$.

thus,

$-Pr(\text{rej } H_0 | H_0) \approx -m[(1-y)\frac{\alpha}{m} + y F_1(\frac{\alpha}{m})]$

$Pr(\text{rej } H_0 | H_0) \approx m[(1-y)\frac{\alpha}{m} + y F_1(\frac{\alpha}{m})]$

(d) For α,

$\frac{\partial Pr(\text{rej } H_0|H_0)}{\partial \alpha} = 1-y + my \frac{\partial F_1(\alpha/m)}{\partial \alpha}$

$= 1-y + my \frac{\partial F_1(\alpha/m)}{\partial \alpha/m} \cdot \frac{\partial \alpha/m}{\partial \alpha} = 1-y + y \frac{\partial F_1(\alpha/m)}{\partial \alpha/m}$

$\because \; 0 < \frac{\partial F_1(\alpha/m)}{\partial \alpha/m} < 1$

$-1 < \frac{\partial F_1(\alpha/m)}{\partial \alpha/m} - 1 < 0$

$1-y < y(\frac{\partial F_1(\alpha/m)}{\partial \alpha/m} - 1) + 1 < 1$

$\because \; 1-y > 0$

$\therefore \; \frac{\partial Pr(\text{rej } H_0 | H_0)}{\partial \alpha} > 0$

thus, with α increasing, prob of rejection increasing.

For y_1

$$\frac{\partial \Pr(\text{rej } H_0 | H_0)}{\partial y} = -\alpha + mF_1(\tfrac{\alpha}{m})$$

thus, how the change of y influence the change of prob of rejection depends on F_1

(e) The bonferonni procedure is applied well in needle in a bag stack scenario, which means that if P_1's shape has sharp peak) and F_1 is not flat, the Bonferonni will have good power as $m \to \infty$

2. See Appendix

3. (1) $L(\theta_j) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}(y_j - \theta_j)^2\}$

$l(\theta_j) = -\frac{1}{2}(y_j - \theta_j)^2 + c$

$\frac{\partial l(\theta_j)}{\partial \theta_j} = y_j - \theta_j = 0$

The MLE for θ_j is $\hat{\theta}_j = y_j$

(2) $H_0: \theta_1 = \theta_2 = \cdots \theta_m = 0$

$-2\log p(y|\theta) = m\log 2\pi + \Sigma(y_j - \theta_j)^2$

$-2\log p(y|\theta_0) = m\log 2\pi + \Sigma y_j^2$

$-2\log p(y|\hat{\theta}) = m\log 2\pi + \Sigma(y_j - \hat{\theta}_j)^2 = m\log 2\pi$

$-2L = \Sigma y_j^2$

level α testing:

choose c so that $Pr(-2L > c \mid H_0) \approx \alpha$

reject H_0 if $-2L > \chi^2_{m, 1-\alpha}$

(3) ① $Pr(p_j < c) = Pr(2 \times \Phi(-|y_i|) < c)$

$= Pr(2 \times (1 - \Phi(|y_i|)) < c)$

$= Pr(\Phi(|y_i|) > 1 - c/2)$

$= Pr(|y_i| > \Phi^{-1}(1 - c/2))$

$= Pr(|y_i| > Z_{1-c/2})$

$= c/2 + c/2 = c$

thus, p_j has a uniform distribution

② $Pr(\text{rej } H_0 | H_0) = pr(\text{rej } H_1 \text{ or rej } H_2 \cdots \text{rej } H_m | H_0)$

$$= Pr(\cup \{ p_j \le \alpha/m \} | H_0)$$

$$\le \sum pr(p_j \le \alpha/m | H_0)$$

$$= \sum \alpha/m$$

$$= \alpha_0$$

thus, this procedure will control the global error rate to be less than or equal to α, even if y_i's are correlated

4. (a) From lecture notes, we have that for $p_j \sim (1-y)P_0 + y P_1$

$$FDP = \frac{\#\{H_j\text{'s rej and equal to zero}\}}{\#\{H_j\text{'s rej}\}}$$

$$= \frac{\sum_{j=1}^{m} 1(p_j < \alpha_E \text{ and } H_j = 0)}{\sum_{j=1}^{m} 1(p_j < \alpha_E)}$$

$$= \frac{\sum D_j(1-H_j)}{D_j}$$, with α_E noted as type I error rate

$$FDR = E(FDP) \overset{\text{tower}}{=} E[\Sigma D_j\, E[(1-H_j) \mid \underset{\sim}{D}] / \Sigma D_j] = E(1-H \mid D=1)$$

$$= Pr(H=0 \mid D=1)$$

$$Pr(H=0 \mid D=1) = Pr(H=0 \mid p < \alpha_E) = \frac{Pr(p < \alpha_E \mid H=0)\, Pr(H=0)}{Pr(p < \alpha_E)}$$

$$= \frac{\alpha_E(1-y)}{F(\alpha_E)}$$

The pdf for beta distribution is

$$P_b(x) = \frac{(1-x)^{b-1}\,\Gamma(b+1)}{\Gamma(1)\,\Gamma(b)} = (1-x)^{b-1}\, b$$

The corresponding cdf is

$$F_b(x) = \int_0^{n} (1-x)^{b-1} \cdot b\, dx = 1-(1-x)^b.$$

The cdf for uniform(0,1) is

$$F_u(x) = x \quad \text{in } (0,1)$$

Thus the mixture cdf is

$$F(x) = (1-y)F_u + yF_b = (1-y)x + y[1-(1-x)^b]$$

To control the FDR at level α, we have

$$FDR = \frac{(1-y)\alpha_E}{(1-y)\alpha_E + y[1-(1-\alpha_E)^b]} = \alpha$$

(b) For mixture distribution

$$E[H(X)] = \sum_{i=1}^{n} w_i \, E[H(x_i)]$$

$$E[(x-\mu)^j] = \sum_{i=1}^{n} w_i \sum_{k=0}^{j} \binom{j}{k} (\mu_i-\mu)^{j-k} E[(x_i-\mu_i)^k]$$

We have the mean and variance of uniform and beta.

Uniform(0,1) : $E(X) = \frac{1}{2}$, $Var(X) = \frac{1}{12}$

beta(1, b) : $E(x) = \frac{1}{1+b}$, $Var(x) = \frac{b}{(1+b)^2(2+b)}$

Therefore, the mean and variance of p_1 are

$$E(p_1) = \sum_{i=1}^{n} w_i \mu_i = (1-y)\frac{1}{2} + y \cdot \frac{1}{1+b}$$

$$Var(p_1) = \sum_{i=1}^{n} w_i(\sigma_i^2 + \mu_i^2) - E(p_1)^2$$

$$= (1-y)(\tfrac{1}{4} + \tfrac{1}{12}) + y\left(\frac{1}{(1+b)^2} + \frac{b}{(1+b)^2(2+b)}\right) - \left[\frac{1-y}{2} + \frac{y}{1+b}\right]^2$$

$$= \frac{1}{12}(1-y)^2 + y^2 \frac{b}{(1+b)^2(b+2)}$$

With observed values of $p_1 \cdots p_m$

$$E(p_1) = \frac{1}{m}\sum_{i=1}^{m} p_i = \bar{p}$$

$$Var(p_1) = \frac{1}{m-1}\sum_{i=1}^{m}(p_i - \bar{p})^2 = s^2$$

With the equation of mean and variance,

$$\bar{p} = \frac{1-y}{2} + \frac{y}{1+b}$$

$$s^2 = \frac{1}{12}(1-y)^2 + y^2 \frac{b}{(1+b)^2(b+2)}$$

STA 532 Homework 10 - Appendix

Yi Mi

Chi-square test

Situation 1

$\theta_1 ... \theta_m \sim$ i.i.d. $N(0, K/100)$ for $K \in \{1, 4, 16, 64\}$.

```
compute_prob.chi = function(K){
  m = 100
  N = 10000
  rej = 0
  for (n in 1:N){
    theta = rnorm(m, 0, sqrt(K/100))
    y = rnorm(m, theta, 1)
    obs = sum(y^2)
    p = pchisq(q = obs, df = m, lower.tail = FALSE)
    if(p<0.05){
      rej = rej + 1
    }
  }
  return(rej/N)
}

chia1 = compute_prob.chi(1)
chia4 = compute_prob.chi(4)
chia16 = compute_prob.chi(16)
chia64 = compute_prob.chi(64)

chi_a = c(chia1, chia4, chia16, chia64)
```

Situation 2

$\theta_1 = K$ and $\theta_2 = ... = \theta_m = 0,$ where $K \in \{1, 3, 5, 7\}$.

```
compute_prob.chi2 = function(K){
  m = 100
  N = 10000
  rej = 0
  for (n in 1:N){
    y = rnorm(m-1, 0, 1)
    theta1 = K
    y1 = rnorm(1, theta1, 1)
    obs = sum(y^2+y1)
    p = pchisq(q = obs, df = m, lower.tail = FALSE)
    if(p<0.05){
      rej = rej + 1
    }
```

```
  }
  return(rej/N)
}

chib1 = compute_prob.chi2(1)
chib3 = compute_prob.chi2(3)
chib5 = compute_prob.chi2(5)
chib7 = compute_prob.chi2(7)
chi_b = c(chib1, chib3, chib5, chib7)
```

Bonferroni's procedure

Situation 1

$\theta_1 ... \theta_m \sim$ i.i.d. $N(0, K/100)$ for $K \in \{1, 4, 16, 64\}$.

```
compute_prob.b = function(K){
  m = 100
  N = 10000
  rej = 0
  for (n in 1:N){
    theta = rnorm(m, 0, sqrt(K/100))
    y = rnorm(m, theta, 1)
    p = 2*(1-pnorm(abs(y)))
    if(any(p<0.05/m)){
      rej = rej + 1
    }
  }
  return(rej/N)
}

bonfa1 = compute_prob.b(1)
bonfa4 = compute_prob.b(4)
bonfa16 = compute_prob.b(16)
bonfa64 = compute_prob.b(64)
bonf_a = c(bonfa1, bonfa4, bonfa16, bonfa64)
```

Situation 2

$\theta_1 = K$ and $\theta_2 = ... = \theta_m = 0,$ where $K \in \{1, 3, 5, 7\}$.

```
compute_prob.b2 = function(K){
  m = 100
  N = 10000
  rej = 0
  for (n in 1:N){
    y = rnorm(m-1, 0, 1)
    theta1 = K
    y[100] = rnorm(1, theta1, 1)
    p = 2*(1-pnorm(abs(y)))
```

```
    if(any(p<0.05/m)){
      rej = rej + 1
    }
  }
  return(rej/N)
}

bonfb1 = compute_prob.b2(1)
bonfb3 = compute_prob.b2(3)
bonfb5 = compute_prob.b2(5)
bonfb7 = compute_prob.b2(7)
bonf_b = c(bonfb1, bonfb3, bonfb5, bonfb7)
```

Fisher's procedure

Situation 1

$\theta_1 ... \theta_m \sim$ i.i.d. $N(0, K/100)$ for $K \in \{1, 4, 16, 64\}$.

```
compute_prob.f = function(K){
  m = 100
  N = 10000
  rej = 0
  for (n in 1:N){
    theta = rnorm(m, 0, sqrt(K/100))
    y = rnorm(m, theta, 1)
    p = 2*(1-pnorm(abs(y)))
    obs = -2*sum(log(p))
    P = pchisq(q = obs, df = 2*m, lower.tail = FALSE)
    if(P<0.05){
      rej = rej + 1
    }
  }
  return(rej/N)
}

fisha1 = compute_prob.f(1)
fisha4 = compute_prob.f(4)
fisha16 = compute_prob.f(16)
fisha64 = compute_prob.f(64)
fish_a = c(fisha1, fisha4, fisha16, fisha64)
```

Situation 2

$\theta_1 = K$ and $\theta_2 = ... = \theta_m = 0,$ where $K \in \{1, 3, 5, 7\}$.

```
compute_prob.f2 = function(K){
  m = 100
  N = 10000
  rej = 0
```

```r
  for (n in 1:N){
    y = rnorm(m-1, 0, 1)
    theta1 = K
    y[100] = rnorm(1, theta1, 1)
    p = 2*(1-pnorm(abs(y)))
    obs = -2*sum(log(p))
    P = pchisq(q = obs, df = 2*m, lower.tail = FALSE)
    if(P<0.05){
      rej = rej + 1
    }
  }
  return(rej/N)
}

fishb1 = compute_prob.f2(1)
fishb3 = compute_prob.f2(3)
fishb5 = compute_prob.f2(5)
fishb7 = compute_prob.f2(7)
fish_b = c(fishb1, fishb3, fishb5, fishb7)
```

Conclusion

```r
df.1 = t(data.frame('Chi-square' = chi_a,
                    'Bonferroni' = bonf_a,
                    'Fisher' = fish_a))
colnames(df.1) = c('K=1', 'K=4', 'K=16', 'K=64')
df.1%>%kable(caption = 'Situation 1 in (a)')
```

Table 1: Situation 1 in (a)

	K=1	K=4	K=16	K=64
Chi.square	0.0550	0.0856	0.2878	0.9666
Bonferroni	0.0531	0.0629	0.1162	0.4844
Fisher	0.0579	0.0859	0.2935	0.9658

```r
df.2 = t(data.frame('Chi-square' = chi_b,
                    'Bonferroni' = bonf_b,
                    'Fisher' = fish_b))
colnames(df.2) = c('K=1', 'K=3', 'K=5', 'K=7')
df.2%>%kable(caption = 'Situation 2 in (b)')
```

Table 2: Situation 2 in (b)

	K=1	K=3	K=5	K=7
Chi.square	0.7660	0.9961	1.0000	1.0000
Bonferroni	0.0542	0.3554	0.9362	0.9998
Fisher	0.0606	0.1348	0.3712	0.7674