Math 181B: Homework 1

Merrick Qiu

Exercise 1

We have that $H_0 = \{\mu_X = \mu_Y \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+\}$ and $\Omega = \{\mu_X, \mu_Y \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+\}$. The likelihood function is

$$\begin{aligned}
L(\theta) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{X_i - \mu_X}{\sigma}\right)^2\right) \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{Y_i - \mu_Y}{\sigma}\right)^2\right) \\
&= (2\pi\sigma^2)^{-\frac{n+m}{2}} \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(X_i - \mu_X)^2 + \sum_{i=1}^{m}(Y_i - \mu_Y)^2\right)\right)
\end{aligned}$$

The log likelihood is

$$\log L(\theta) = -\frac{n+m}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(X_i - \mu_X)^2 + \sum_{i=1}^{m}(Y_i - \mu_Y)^2\right)$$

Under H_0 we have that

$$\begin{aligned}
\frac{\partial}{\partial\mu}\log L(\theta) = \frac{1}{\sigma^2}\left(\sum_{i=1}^{n}(X_i - \mu) + \sum_{i=1}^{m}(Y_i - \mu)\right) &= 0 \\
\implies \sum_{i=1}^{n} X_i + \sum_{i=1}^{m} Y_i &= (n+m)\hat{\mu} \\
\implies \hat{\mu} &= \frac{n\bar{X} + m\bar{Y}}{n+m}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial\sigma^2}\log L(\theta) = -\frac{n+m}{2\sigma^2} + \frac{1}{2\sigma^4}\left(\sum_{i=1}^{n}(X_i - \mu)^2 + \sum_{i=1}^{m}(Y_i - \mu)^2\right) &= 0 \\
\implies \left(\sum_{i=1}^{n}(X_i - \mu)^2 + \sum_{i=1}^{m}(Y_i - \mu)^2\right) &= (n+m)\hat{\sigma}^2 \\
\implies \hat{\sigma}^2 = \frac{1}{n+m}\left(\sum_{i=1}^{n}(X_i - \hat{\mu})^2 + \sum_{i=1}^{m}(Y_i - \hat{\mu})^2\right)
\end{aligned}$$

$$\begin{aligned}
\max_{\theta\in H_0} L(\theta) &= (2\pi\hat{\sigma}^2)^{-\frac{n+m}{2}} \exp\left(-\frac{1}{2\hat{\sigma}^2}\left(\sum_{i=1}^{n}(X_i - \hat{\mu})^2 + \sum_{i=1}^{m}(Y_i - \hat{\mu})^2\right)\right) \\
&= \left(\frac{2\pi}{n+m}\left(\sum_{i=1}^{n}(X_i - \hat{\mu})^2 + \sum_{i=1}^{m}(Y_i - \hat{\mu})^2\right)\right)^{-\frac{n+m}{2}} \exp\left(-\frac{n+m}{2}\right)
\end{aligned}$$

Under Ω we have that

$$\begin{aligned}\frac{\partial}{\partial \mu_x} \log L(\theta) &= \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu_X) = 0 \\ \implies n\hat{\mu}_X &= \frac{1}{\sigma^2} \sum_{i=1}^{n} X_i \\ \implies \hat{\mu}_X &= \bar{X}\end{aligned}$$

Since its symmetric we also have $\hat{\mu}_y = \bar{Y}$.

$$\begin{aligned}\frac{\partial}{\partial \sigma^2} \log L(\theta) = -\frac{n+m}{2\sigma^2} + \frac{1}{2\sigma^4} \left(\sum_{i=1}^{n} (X_i - \mu_X)^2 + \sum_{i=1}^{m} (Y_i - \mu_Y)^2 \right) = 0 \\ \implies \left(\sum_{i=1}^{n} (X_i - \bar{X})^2 + \sum_{i=1}^{m} (Y_i - \bar{Y})^2 \right) = (n+m)\hat{\sigma}^2 \\ \implies \hat{\sigma}^2 = \frac{1}{n+m} \left(\sum_{i=1}^{n} (X_i - \bar{X})^2 + \sum_{i=1}^{m} (Y_i - \bar{Y})^2 \right)\end{aligned}$$

$$\begin{aligned}\max_{\theta \in \Omega} L(\theta) &= (2\pi\hat{\sigma}^2)^{-\frac{n+m}{2}} \exp\left(-\frac{1}{2\hat{\sigma}^2} \left(\sum_{i=1}^{n} (X_i - \bar{X})^2 + \sum_{i=1}^{m} (Y_i - \bar{Y})^2 \right) \right) \\ &= \left(\frac{2\pi}{n+m} \left(\sum_{i=1}^{n} (X_i - \bar{X})^2 + \sum_{i=1}^{m} (Y_i - \bar{Y})^2 \right) \right)^{-\frac{n+m}{2}} \exp\left(-\frac{n+m}{2} \right)\end{aligned}$$

The likelihood ratio is

$$\begin{aligned}\Lambda &= \frac{\left(\sum_{i=1}^{n} (X_i - \bar{X})^2 + \sum_{i=1}^{m} (Y_i - \bar{Y})^2\right)^{-\frac{n+m}{2}}}{\left(\sum_{i=1}^{n} (X_i - \hat{\mu})^2 + \sum_{i=1}^{m} (Y_i - \hat{\mu})^2\right)^{-\frac{n+m}{2}}} \\ &= \left(\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2 + n(\bar{X} - \hat{\mu})^2 + \sum_{i=1}^{m} (Y_i - \hat{\mu})^2 + m(\bar{Y} - \hat{\mu})^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2 + \sum_{i=1}^{m} (Y_i - \bar{Y})^2} \right)^{\frac{n+m}{2}} \\ &= \left(1 + \frac{n\left(\frac{m}{n+m}(\bar{X} - \bar{Y})\right)^2 + m\left(-\frac{n}{n+m}(\bar{X} - \bar{Y})\right)^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2 + \sum_{i=1}^{m} (Y_i - \bar{Y})^2} \right)^{\frac{n+m}{2}} \\ &= \left(1 + \frac{nm}{n+m} \frac{(\bar{X} - \bar{Y})^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2 + \sum_{i=1}^{m} (Y_i - \bar{Y})^2} \right)^{\frac{n+m}{2}}\end{aligned}$$

For some c, the test is rejects when

$$\left(1 + \frac{nm}{n+m} \frac{(\bar{X} - \bar{Y})^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2 + \sum_{i=1}^{m} (Y_i - \bar{Y})^2} \right)^{\frac{n+m}{2}} \le c$$

We can rewrite this as

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 + \sum_{i=1}^{m}(Y_i - \bar{Y})^2}} \leq c'$$

for some c', which is equivalent to the two sample t-test.

Exercise 2

1. ```
   # Do a HT on H0: mu_X = mu_Y and H1: mu_X > mu_Y.
   # We cannot assume that sigma_x = sigma_y,
   # so we use Welch's approximation for this calculation.

   # Import files
   setwd("C:/Users/merri/Documents/MATH-31H/MATH 181B/Homework 1")
   phone = unlist(read.csv("Phone.csv"))
   noPhone = unlist(read.csv("NoPhone.csv"))

   # Calculate mean and std
   Xbar = mean(phone)
   Ybar = mean(noPhone)
   Sx = sd(unlist(phone))
   Sy = sd(unlist(noPhone))
   n = length(phone)
   m = length(noPhone)

   # The test statistic is
   Tv = (Xbar - Ybar)/sqrt(Sx^2/n + Sy^2/m)

   # v degrees of freedom
   v = round((Sx^2/n + Sy^2/m)^2/(Sx^4/n^2/(n-1) + Sy^4/m^2/(m-1)))

   # Find P(t_56 > 2.44)
   pt(Tv, v, lower=F)

   # Since p = 0.008970693 < 0.05, we reject the null hypothesis
   ```

2. ```
   # Calculate Confidence interval
   diffMean = Xbar - Ybar
   ME = qt(0.005, df = v, lower.tail=FALSE)*sqrt(Sx^2/n + Sy^2/m)
   cat("(", diffMean - ME, ", ", diffMean + ME, ")")
   # The confidence interval is ( -2.976387 , 66.66539 )
   ```

   99% of confidence intervals have the mean in it. We reject if $\mu_X - \mu_y = 0$ is not in the confidence interval. Since 0 is in the interval, we fail to reject, which is in contrast to a).

3. ```
   # Verify
   t.test(phone, noPhone, alternative = "greater")$p.value
   t.test(phone, noPhone, conf.level = 0.99)$conf.int
   # Output
   # 0.008956362
   # -2.965777 66.654780
   ```
   ```

Exercise 3

The pooled variance is

$$S_p^2 = \frac{1^2 + 1.5^2}{2} = 1.625$$

The margin of error is

$$ME = t_{0.005,18}\sqrt{1.625}\sqrt{1/10 + 1/10} = 1.640964$$

Therefore the values that ? can take on are in the confidence interval (3.36, 6.64).