STA200C HW1

Draft version 1. Some notations to be corrected (like bold, arrows, superscripts, subscripts). Some steps to be filled. Feel free to discuss in office hours if any confusing spots. Chi Po Choi.

1. Let $S = \sum_{i=1}^{m} Y_i$. The log likelihood function is

$$L(\theta|Y_1,\ldots,Y_m) = \theta^S(1-\theta)^{m-S}$$

   The MLE is $\min\{\frac{S}{m}, \theta_0\}$ in H_0 and $\frac{S}{m}$ in H_1.

   The LRT statistics is

$$\begin{aligned}\lambda(\mathbf{Y}) &= \frac{\sup_{\theta\leq\theta_0} L(\theta|Y_1,\ldots,Y_m)}{\sup_{\theta\in\mathbb{R}} L(\theta|Y_1,\ldots,Y_m)} \\ &= \begin{cases} 1 & \text{if } \frac{S}{m} \leq \theta_0 \\ \frac{(\theta_0)^S(1-\theta_0)^{m-S}}{(\frac{S}{m})^S(1-\frac{S}{m})^{m-S}} & \text{if } \frac{S}{m} > \theta_0 \end{cases}\end{aligned}$$

   The LRT is: H_0 will be rejected if

$$\frac{(\theta_0)^S(1-\theta_0)^{m-S}}{(\frac{S}{m})^S(1-\frac{S}{m})^{m-S}} < c \qquad \text{where } c \text{ is the critical value}$$

   Do some calculus and show $\log\lambda(S)$ is a decreasing function in S. Then the test is equivalent to $S > b$ for some constant b.

2. (a) Hypothesis: $H_0 : \theta = \mu$ versus $H_1 : \theta \neq \mu$

   Write down the likelihood function of exponential variable sample. Do calculus to find the MLE in H_0 and H_1. The MLE is $\hat\theta = (\sum x + \sum y)/(n+m)$ under H_0. The MLE is $\hat\theta = \bar{x}$, $\hat\theta = \bar{y}$ under H_1. Put the MLEs into the likelihood functions.

   Likelihood Ratio Test is:

   H_0 will be rejected if $\lambda(\mathbf{X},\mathbf{Y}) < c$ where

$$\lambda(\mathbf{X},\mathbf{Y}) = \frac{(n+m)^{n+m}}{n^n m^m} \frac{(\sum_i X_i)^n \left(\sum_j Y_j\right)^m}{\left(\sum_i X_i + \sum_j Y_j\right)^{n+m}}$$

   and c is the critical value.

   Then show the polynomial in T above is a concave/convex function.

(b) Let $T = \frac{\sum X_i}{\sum X_i + \sum Y_i}$. Then

$$\begin{aligned}
\lambda(\mathbf{X}, \mathbf{Y}) &= \frac{(n+m)^{n+m}}{n^n m^m} \frac{\left(\sum_i X_i\right)^n \left(\sum_j Y_j\right)^m}{\left(\sum_i X_i + \sum_j Y_j\right)^{n+m}} \\
&= \frac{(n+m)^{n+m}}{n^n m^m} \left(\frac{\sum_i X_i}{\sum_i X_i + \sum_j Y_j}\right)^n \left(\frac{\sum_j Y_j}{\sum_i X_i + \sum_j Y_j}\right)^m \\
&= \frac{(n+m)^{n+m}}{n^n m^m} T^n (1-T)^m
\end{aligned}$$

(c) When H_0 is true, a.k.a. $\theta = \mu$,

$$\text{If } X_1, \ldots X_n \sim \text{exponential}(\theta) \text{ independently then } \sum_i X_i \sim \text{Gamma}(n, \theta)$$

$$\text{If } Y_1, \ldots Y_m \sim \text{exponential}(\theta) \text{ independently then } \sum_j Y_j \sim \text{Gamma}(m, \theta)$$

Therefore,

$$T = \frac{\sum_i X_i}{\sum_i X_i + \sum_j Y_j} \sim \text{Beta(n, m)}$$

Reference
`https://en.wikipedia.org/wiki/Exponential_distribution#Related_distributions`
`https://en.wikipedia.org/wiki/Gamma_distribution#General`

3. Model

$$y_{i,j} = \mu_j + \varepsilon_{i,j}$$

where

$i = 1, \ldots, I$ is an index over experimental units;

$j = 1, \ldots, J$ is an index over treatment groups;

I_j is the number of experimental units in the jth treatment group;

$I = \sum_j I_j$ is the total number of experimental units;

$y_{i,j}$ are observations;

μ_j is the mean of the observations for the jth treatment group;

$\varepsilon \sim N(0, \sigma^2)$, $\varepsilon_{i,j}$ are independently normally distributed zero-mean random errors.

The hypothesis is $H_0 : \mu_j$'s are all equal vs $H_1 : \mu_j$'s are not all equal.

The log likelihood function of the model is

$$\log L(\mu_j, \sigma^2 | y_{i,j}) = -\frac{\sum_j I_j}{2}(\log 2\pi + \log \sigma^2) - \frac{1}{2\sigma^2}\left(\sum_j \sum_{i=1}^{I_j} (y_{i,j} - \mu_j)^2\right).$$

Write down the MLEs and put them into the likelihood functions.

The sup log likelihood function under H_0 is

$$\sup_{\theta \in \Theta_0} \log L(\mu_j, \sigma^2 | y_{i,j}) = -\frac{\sum_j I_j}{2}(\log 2\pi + \log \hat{\sigma}_0^2) - \frac{\sum_j I_j}{2}$$

where

$$\hat{\sigma}_0^2 = \frac{1}{\sum_j I_j} \sum_j \sum_{i=1}^{I_j} (y_{i,j} - \bar{y}_{\cdot,\cdot})^2 \qquad \bar{y}_{\cdot,\cdot} = \frac{\sum_j \sum_{i=1}^{I_j} y_{i,j}}{\sum_j I_j}.$$

The sup log likelihood function over whole parameter space is

$$\sup_{\theta \in \Theta} \log L(\mu_j, \sigma^2 | y_{i,j}) = -\frac{\sum_j I_j}{2}(\log 2\pi + \log \hat{\sigma}^2) - \frac{\sum_j I_j}{2}$$

where

$$\hat{\sigma}^2 = \frac{1}{\sum_j I_j} \sum_j \sum_{i=1}^{I_j} (y_{i,j} - \bar{y}_{\cdot,j})^2 \qquad \bar{y}_{\cdot,j} = \frac{\sum_{i=1}^{I_j} y_{i,j}}{I_j}.$$

Therefore, the log likelihood ratio is

$$\log \lambda(y_{i,j}) = -\frac{\sum_j I_j}{2}(\log \hat{\sigma}_0^2 - \log \hat{\sigma}^2) = -\frac{\sum_j I_j}{2} \log \frac{\hat{\sigma}_0^2}{\hat{\sigma}^2}$$

The LRT is: reject H_0 if $\lambda(y_{i,j}) < c$ for some critical values.

4. Similar to Question 3, but $J = 2$ and one-sided.

$$\log L(\mu_1, \mu_2, \sigma^2 | X_{1j}\text{'s}, X_{2j}\text{'s}) = -\frac{n_1 + n_2}{2}(\log 2\pi + \log \sigma^2) - \frac{1}{2\sigma^2}\left(\sum_j \sum_{i=1}^{n_j} (X_{ij} - \mu_j)^2\right).$$

The sup log likelihood function under H_0 is

$$\sup_{\theta \in \Theta_0} \log L(\mu_j, \sigma^2 | X_{ij}) = \begin{cases} -\frac{n_1+n_2}{2}(\log 2\pi + \log \hat{\sigma}_0^2) - \frac{n_1+n_2}{2} & \text{if } \overline{X}_{1\cdot} > \overline{X}_{2\cdot} \\ -\frac{n_1+n_2}{2}(\log 2\pi + \log \hat{\sigma}^2) - \frac{n_1+n_2}{2} & \text{if } \overline{X}_{1\cdot} \leq \overline{X}_{2\cdot} \end{cases}$$

where

$$\hat{\sigma}_0^2 = \frac{1}{\sum_j n_j} \sum_j \sum_{i=1}^{n_j} \left(X_{ij} - \overline{X}_{\cdot\cdot}\right)^2 \qquad \hat{\sigma}^2 = \frac{1}{\sum_j n_j} \sum_j \sum_{i=1}^{n_j} \left(X_{ij} - \overline{X}_{\cdot j}\right)^2$$

$$\overline{X}_{\cdot j} = \frac{\sum_{i=1}^{n_j} X_{ij}}{n_j} \qquad \overline{X}_{\cdot\cdot} = \frac{\sum_j \sum_{i=1}^{n_j} X_{ij}}{\sum_j n_j}$$

The sup log likelihood function in general is

$$\sup_{\theta\in\Theta} \log L(\mu_j, \sigma^2 | X_{ij}) = -\frac{n_1+n_2}{2}(\log 2\pi + \log \hat{\sigma}^2) - \frac{n_1+n_2}{2}$$

Therefore, the log likelihood ratio is

$$\log \lambda(X_{ij}) = \begin{cases} -\frac{n_1+n_2}{2} \log \frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} & \text{if } \overline{X}_{1\cdot} > \overline{X}_{2\cdot} \\ 1 & \text{if } \overline{X}_{1\cdot} \leq \overline{X}_{2\cdot} \end{cases}$$

The LRT is: H_0 will be rejected if $\lambda < c$ for some critical values c.

To show it is equivalent to two sample t test, rewrite the expression into:

$$\begin{aligned} \frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} &= \frac{S_1 + S_2 + n_1 n_2 (\overline{X}_{\cdot 1} - \overline{X}_{\cdot 2})^2}{S_1 + S_2} \\ &= 1 + n_1 n_2 \frac{(\overline{X}_{\cdot 1} - \overline{X}_{\cdot 2})^2}{S_1 + S_2} \end{aligned}$$

where $S_j = \sum_i (X_{ij} - \overline{X}_{\cdot j})^2$. It is a monotone function in

$$t = \frac{\overline{X}_{\cdot 1} - \overline{X}_{\cdot 2}}{\sqrt{\frac{S_1+S_2}{n_1+n_2-2}}\sqrt{1/n_1 + 1/n_2}}.$$

5. The negative log likelihood function is

$$-\log L(\vec{\mu}, \mathbf{\Sigma} | \vec{X}) = \frac{n}{2} \log \det(\mathbf{\Sigma}) + \frac{1}{2} \sum_{i=1}^{n} (\vec{X}_i - \vec{\mu})^T \mathbf{\Sigma}^{-1} (\vec{X}_i - \vec{\mu}) + \frac{n}{2} \log 2\pi$$

The sup negative log likelihood function under H_0 is

$$\begin{aligned} \sup_{\theta\in\Theta_0} -\log L(\vec{\mu}, \mathbf{\Sigma} | \vec{X}) &= \frac{n}{2} \log \det(\mathbf{S}_0) + \frac{1}{2} \sum_{i=1}^{n} (\vec{X}_i - \vec{\mu}_0)^T \mathbf{S}_0^{-1} (\vec{X}_i - \vec{\mu}_0) + \frac{n}{2} \log 2\pi \\ &= \frac{n}{2} \log \det(\mathbf{S}_0) + \frac{np}{2} + \frac{n}{2} \log 2\pi \end{aligned}$$

where

$$\mathbf{S}_0 = \frac{1}{n} \sum_{i=1}^{n} (\vec{X}_i - \vec{\mu}_0)(\vec{X}_i - \vec{\mu}_0)^T$$

The sup negative log likelihood function over whole parameter space is

$$\begin{aligned}
\sup_{\theta\in\Theta} -\log L(\vec{\mu},\mathbf{\Sigma}|\vec{X}) &= \frac{n}{2}\log\det(\mathbf{S}) + \frac{1}{2}\sum_{i=1}^{n}(\vec{X}_i-\overline{X})^T\mathbf{S}^{-1}(\vec{X}_i-\overline{X}) + \frac{n}{2}\log 2\pi \\
&= \frac{n}{2}\log\det(\mathbf{S}) + \frac{np}{2} + \frac{n}{2}\log 2\pi
\end{aligned}$$

where

$$\mathbf{S} = \frac{1}{n}\sum_{i=1}^{n}(\vec{X}_i-\overline{X})(\vec{X}_i-\overline{X})^T \qquad \overline{X} = \frac{1}{n}\sum_{i=1}^{n}\vec{X}_i$$

The log likelihood ratio is

$$\log\lambda = \frac{n}{2}(\log\det\mathbf{S} - \log\det\mathbf{S}_0)$$

The LRT is: H_0 will be rejected if $\lambda < c$ for some critical values c.

To show the LRT is equivalent to Hotelling's t-squared, rewrite the expression into

$$\begin{aligned}
nS_0 &= S + n(\overline{X}-\mu_0)(\overline{X}-\mu_0)^T \\
\det(S)/\det(S_0) &= \det(S)/\det(S + n(\overline{X}-\mu_0)(\overline{X}-\mu_0)^T) \\
&= 1/(1 + (\overline{X}-\mu_0)^T S^{-1}(\overline{X}-\mu_0))
\end{aligned}$$

by `https://en.wikipedia.org/wiki/Matrix_determinant_lemma`

$$T^2 = n(\overline{X}-\vec{\mu}_0)^T S^{-1}(\overline{X}-\vec{\mu}_0)$$

$$\lambda^{2/n} = \frac{1}{1+\frac{T^2}{n-1}}$$

Since λ is a monotone function of T^2, they define the same family of rejection region, thus define equivalent tests.

6.

$$H_0 : \theta \in \bigcap_{\vec{a}}\{\vec{\mu}|\vec{a}^T(\vec{\mu}-\vec{\mu}_0)=0\}$$

Let $y = \vec{a}^T\vec{X} \sim N(\vec{a}^T\vec{\mu}, \vec{a}^T\mathbf{\Sigma}\vec{a})$. The problem is reduced to test $H_0 : \vec{a}^T\vec{\mu} = \vec{a}^T\vec{\mu}_0$ at given $\vec{a}$. The rejection region is is

$$\{\mathbf{y} : \left(\frac{\sum_{i=1}^n (y_i - \vec{a}^T\vec{\mu}_0)^2}{\sum_{i=1}^n (y_i-\bar{y})^2}\right)^{-n/2} \le c\}$$

Then

$$\bigcup_{\vec{a}} \{\mathbf{y} : \left(\frac{\sum_{i=1}^n (y_i - \vec{a}^T \vec{\mu}_0)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}\right)^{-n/2} \le c\} = \{\mathbf{y} : \inf_{\vec{a}} \left(\frac{\sum_{i=1}^n (y_i - \vec{a}^T \vec{\mu}_0)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}\right)^{-n/2} \le c\}$$
$$= \{\mathbf{y} : \sup_{\vec{a}} \left(\frac{\sum_{i=1}^n (y_i - \vec{a}^T \vec{\mu}_0)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}\right) \ge c^{-2/n}\}$$

$$\begin{aligned}
\sum_i (y_i - \vec{a}^T \vec{\mu}_0)^2 &= \sum_i (y_i - \bar{y})^2 + n(\bar{y} - \vec{a}^T \vec{\mu}_0)^2 \\
\frac{\sum_{i=1}^n (y_i - \vec{a}^T \vec{\mu}_0)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} &= 1 + \frac{(\bar{y} - \vec{a}^T \vec{\mu}_0)^2}{\frac{1}{n}\sum_{i=1}^n (y_i - \bar{y})^2} \\
&= 1 + \frac{a^T(\bar{x} - \vec{\mu}_0)(\bar{x} - \vec{\mu}_0)^T a}{a^T \mathbf{S} a} \\
&\le 1 + (\bar{x} - \vec{\mu}_0)^T \mathbf{S}^{-1} (\bar{x} - \vec{\mu}_0) = 1 + T^2
\end{aligned}$$

with maximum attained at $a = \sqrt{\mathbf{S}^{-1}}(\bar{x} - \vec{\mu}_0)$ by `https://en.wikipedia.org/wiki/Cauchy%E2%80%93Schwarz_inequality` ($\sqrt{\mathbf{S}^{-1}}$ could be the Cholesky decomposition of $\mathbf{S}^{-1}$, for example.)