Stat601(Section001): Homework #1

Due on Jan. 19, 2021 at 11:59pm

Instructor:Ziwei Zhu

Tiejin Chen **tiejin@umich.edu**

Problem 1

(a)

$\Rightarrow$ Assumed X_1 be $(X_{11},..,X_{1n})$,n-dim vector and X_2 be $(X_{21},..,X_{2m})$, m-dim vector. And we have X_1 and X_2 are independent. That is to say for any $0 < i \leq n, 0 < j \leq m$, X_{1i} and X_{2j} are independent. Now We use σ_{ij} to present i-th row j-th column element of Σ_{12}. We know, for any $0 < i \leq n, 0 < j \leq m$,$\sigma_{ij} = Cov(X_{1i}, X_{2j}) = 0$. Hence $\Sigma_{12} = \Sigma_{21} = 0$.

$\Leftarrow$ if we have $\Sigma_{21} = \Sigma_{12} = 0$, then for Partition$(X_1, X_2)$, we have the Σ_{new}: $\begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{bmatrix}$

Consider the kernel of term in expontional function of the joint pdf of (X_1, X_2). We get:

$$(X_1 - \mu_1, X_2 - \mu_2)^T \begin{bmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{bmatrix} (X_1 - \mu_1, X_2 - \mu_2) = (X_1 - \mu_1)^T \Sigma_{11}^{-1}(X_1 - \mu_1) + (X_2 - \mu_2)^T \Sigma_{22}^{-1}(X_2 - \mu_2)$$

Now we can rewrite the joint pdf of Partition(X_1, X_2), to get:

$$\begin{aligned} f(x) &= \frac{1}{\sqrt{0.5|\Sigma_{new}|}} exp[-0.5((X_1 - \mu_1, X_2 - \mu_2)^T \begin{bmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{bmatrix} (X_1 - \mu_1, X_2 - \mu_2))] \\ &= \frac{1}{\sqrt{0.5|\Sigma_{11}||\Sigma_{22}|}} exp[-0.5((X_1 - \mu_1)^T \Sigma_{11}^{-1}(X_1 - \mu_1))] \times exp[-0.5(X_2 - \mu_2)^T \Sigma_{22}^{-1}(X_2 - \mu_2)] \end{aligned}$$

And we can see that this is mulitply of two mulitvariant normal distributions. the joint pdf can be divided into two pdf of mulitvariant normal distribution. Hence X_1 and X_2 are independent.

b

We know that the distribution of (X_1, X_2) given X_3 is a normal distribution and we use $\Sigma_{12|3}$ to present the conditional covarience matrix of (X_1, X_2) given X_3. We know that:

$$\begin{aligned} \Sigma_{12|3} &= \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} - \begin{bmatrix} \Sigma_{13} \\ \Sigma_{23} \end{bmatrix} \Sigma_{33}^{-1} \begin{bmatrix} \Sigma_{31} \\ \Sigma_{32} \end{bmatrix} \\ &= \begin{bmatrix} \Sigma_{11} - \Sigma_{13}\Sigma_{33}^{-1}\Sigma_{31} & \Sigma_{12} - \Sigma_{13}\Sigma_{33}^{-1}\Sigma_{32} \\ \Sigma_{21} - \Sigma_{23}\Sigma_{33}^{-1}\Sigma_{31} & \Sigma_{22} - \Sigma_{23}\Sigma_{33}^{-1}\Sigma_{32} \end{bmatrix} \end{aligned}$$

Now we only need to prove when $\Sigma_{21} - \Sigma_{23}\Sigma_{33}^{-1}\Sigma_{31} = 0$, $\Lambda_{12} = 0$.We re-partition the Σ to get:

$$\Sigma = \begin{bmatrix} \widehat{\Sigma_{11}} & \widehat{\Sigma_{12}} \\ \widehat{\Sigma_{21}} & \widehat{\Sigma_{22}} \end{bmatrix}$$

where $\widehat{\Sigma_{22}} = \Sigma_{33}, \widehat{\Sigma_{11}} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$. For this block matrix, we can simplifer it to $\begin{bmatrix} \widehat{\Sigma_{11.2}} & 0 \\ 0 & ** \end{bmatrix}$. where ** is a result we do not care. And we have:

$$\Sigma^{-1} = \Lambda = \begin{bmatrix} \widehat{\Sigma_{11.2}}^{-1} & 0 \\ 0 & **^{-1} \end{bmatrix}$$

Hence, we can know:

$$\begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix} = \widehat{\Sigma_{11.2}}^{-1}$$

$$\begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix}^{-1} = \widehat{\Sigma_{11.2}} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} - \begin{bmatrix} \Sigma_{13} \\ \Sigma_{23} \end{bmatrix} \Sigma_{33}^{-1} \begin{bmatrix} \Sigma_{31} \\ \Sigma_{32} \end{bmatrix} = \begin{bmatrix} \Sigma_{11} - \Sigma_{13}\Sigma_{33}^{-1}\Sigma_{31} & \Sigma_{12} - \Sigma_{13}\Sigma_{33}^{-1}\Sigma_{32} \\ \Sigma_{21} - \Sigma_{23}\Sigma_{33}^{-1}\Sigma_{31} & \Sigma_{22} - \Sigma_{23}\Sigma_{33}^{-1}\Sigma_{32} \end{bmatrix}$$

Hence $\Sigma_{21} - \Sigma_{23}\Sigma_{33}^{-1}\Sigma_{31} = 0$ is equivalence to $\Lambda_{12} = 0$, which is the end of our proof.

Problem 2

(a)

The row and column vectors are a little bit confusing here. Hence we stipulate that X_i is a row vector. We only consider the kernel of multivariant normal distribution. We should have:

$$Y \sim N(XB, \Sigma)$$

Hence the kernel of pdf of Y is:

$$f(Y) = |\Sigma|^{-\frac{1}{2}} exp(-\frac{1}{2}(Y - XB)\Sigma^{-1}(Y - XB)^T$$

And the likelihood function of data(with only kernel) will be:

$$L(B, \Sigma|X, Y) = |\Sigma|^{-\frac{N}{2}} exp(-\frac{1}{2}\sum_{i=1}^{N}(Y_i - X_iB)\Sigma^{-1}(Y_i - X_iB)^T)$$

Hence the kerner of log-likelihood function is:

$$l(B, \Sigma|X, Y) = -\frac{N}{2}log(|\Sigma|) - \frac{1}{2}\sum_{i=1}^{N}(Y_i - X_iB)\Sigma^{-1}(Y_i - X_iB)^T$$

Similar to lecture,since $(Y_i - X_iB)^T\Sigma^{-1}(Y_i - X_iB)$ is a scalar, we can use trace here. And we can get:

$$l(B, \Sigma|X, Y) = -\frac{N}{2}log(|\Sigma|) - \frac{1}{2}\sum_{i=1}^{N}tr(\Sigma^{-1}\Sigma_B)$$

And we have:

$$\frac{\partial l(B, \Sigma|X, Y)}{\partial B} = \sum_{i=1}^{N} X_i^T(Y_i - X_iB)\Sigma^{-1} = 0$$

We can get:

$$X^TY = X^TXB \rightarrow \hat{B} = (X^TX)^{-1}X^TY$$

For Σ we have:

$$\frac{\partial l(B, \Sigma|X, Y)}{\partial \Sigma^{-1}} = \frac{N}{2}\Sigma - \frac{1}{2}\sum_{i=1}^{N}(Y_i - X_iB)^T(Y_i - X_iB)$$

Hence:

$$\hat{\Sigma} = \frac{1}{N}\sum_{i=1}^{N}(Y_i - X_iB)^T(Y_i - X_iB)$$

(b)

in this part, X_i and Y_i are column vectors.We have:

$$\frac{\partial RSS(B)}{\partial B} = \sum_{i=1}^{N} -2(Y_i^T - X_i^TB)^TX_i^T = 0$$

We can get:

$$X^TY = X^TXB \rightarrow \hat{B}_{OLS} = (X^TX)^{-1}X^TY$$

We can see it is the same as MLE estimator.

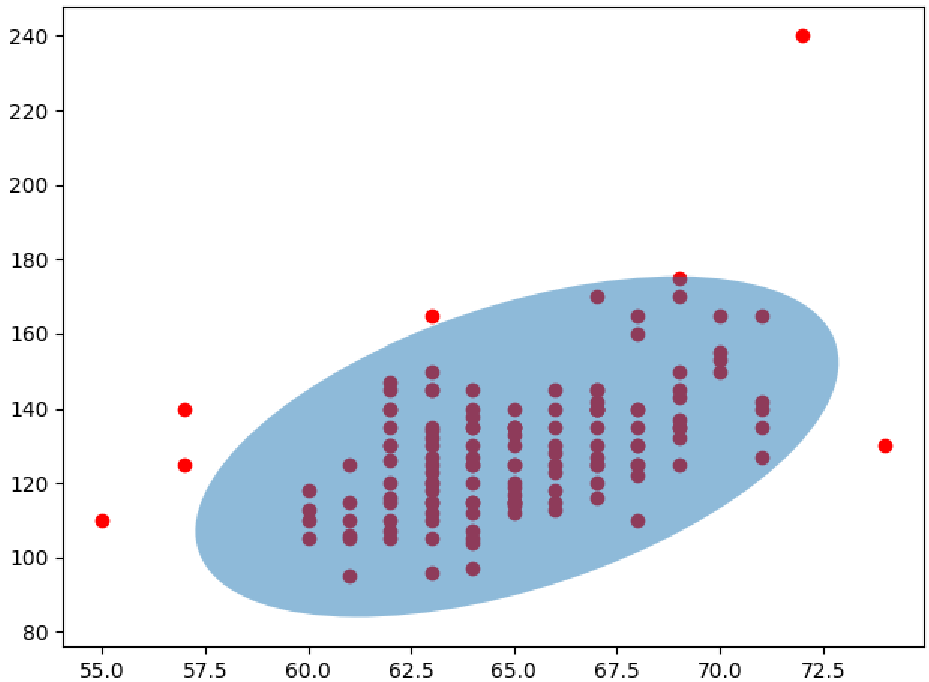
Problem 3

(a)

we find the mean vector is $(65.0729927, 129.69343066)$. And the covariance matrix is:

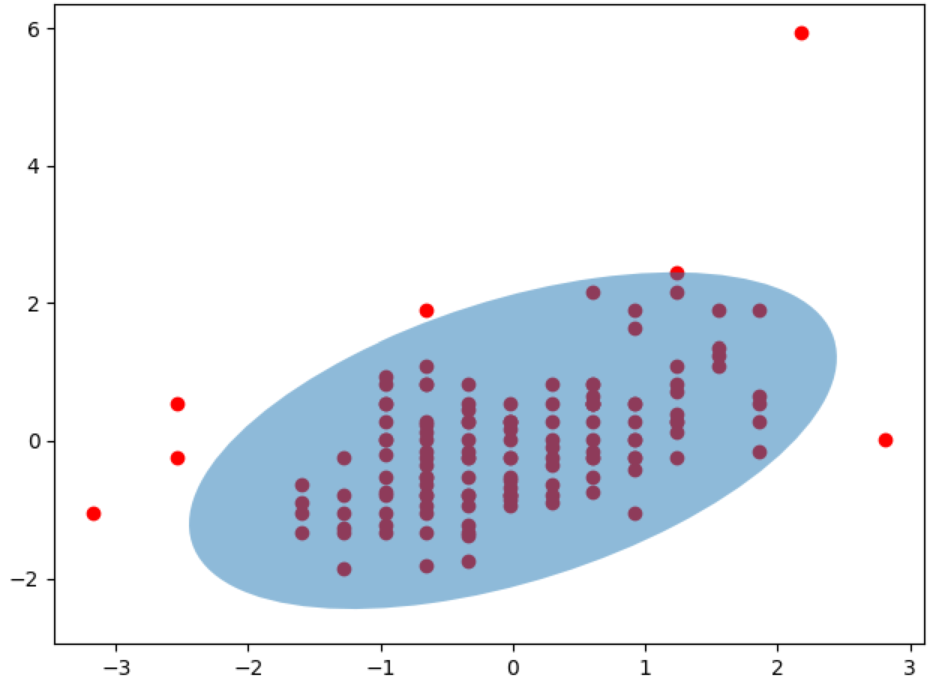
$$\begin{bmatrix} 10.18580936 & 29.53724775 \\ 29.53724775 & 348.53767711 \end{bmatrix}$$

And we get the plot:

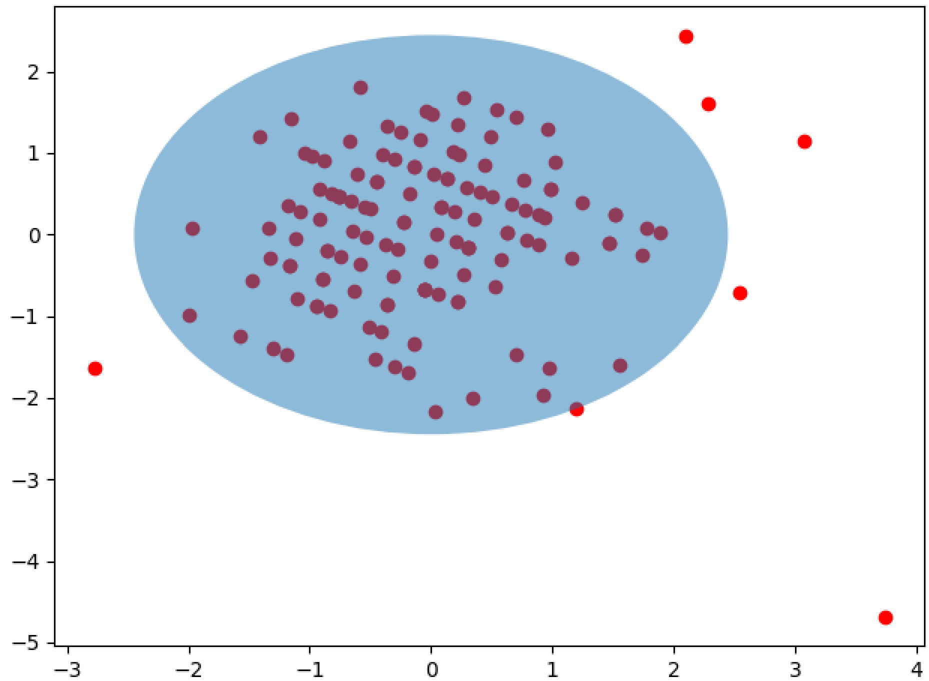


Problem 3 continued on next page...

(b)



(c)



due to the different scale x-axis and y-axis use, it looks like a ellipse. However it is more like a round indeed.

Problem 3 continued on next page. . .

(d)

p-value we get using Hotelling's T test is 1.11e-16(the minimal value in default float in python I guess) which is smaller than 0.05. Hence, we reject H_0 and accept $H_A : \mu_0 \neq \mu_1$

Problem 4

(a)

We also only consider the kernel of pdf here. Hence, we have:

$$LR = \frac{sup_{\Sigma}(|\Sigma|^{-\frac{N}{2}} exp(-0.5\sum_{i=1}^{N} x_i^T \Sigma^{-1} x_i))}{sup_{\mu,\Sigma}(|\Sigma|^{-\frac{N}{2}} exp(-0.5\sum_{i=1}^{N}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)))}$$

Using MLE, we know the denominator becomes max when $\mu = \bar{x_n}$ and $\Sigma = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)(x_i-\mu)^T$. Now we also use MLE to get the MLE estimator of nominator. Very similar to the general form, we only need to use x_i to place $(x_i - \mu)$. And we get the MLE of nominator:

$$\Sigma_{nomi} = \frac{1}{N}\sum_{i=1}^{N} x_i x_i^T$$

Hence we can get:

$$\begin{aligned}
LR &= \frac{|\Sigma_{nomi}|^{-\frac{N}{2}} exp(-0.5tr(|\Sigma_{nomi}|^{-1}\sum_{i=1}^{N} x_i x_i^T))}{|\Sigma_{denomi}|^{-\frac{N}{2}} exp(-0.5tr(|\Sigma_{denomi}|^{-1}\sum_{i=1}^{N}(x_i-\mu)(x_i-\mu)^T))} \\
&= (\frac{|\Sigma_{nomi}|}{|\Sigma_{denomi}|})^{-\frac{N}{2}} \frac{exp(-0.5N(\frac{1}{N}\sum_{i=1}^{N} x_i x_i^T)^{-1}(\frac{1}{N}\sum_{i=1}^{N} x_i x_i^T))}{exp(-0.5N(\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)(x_i-\mu)^T)^{-1}(\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)(x_i-\mu)^T))} \\
&= (\frac{|\Sigma_{nomi}|}{|\Sigma_{denomi}|})^{-\frac{N}{2}}
\end{aligned}$$

for Σ_{nomi}, we have:

$$\begin{aligned}
\sum_{i=1}^{N} x_i x_i^T &= \sum_{i=1}^{N}(x_i - \bar{x_n} + \bar{x_n})(x_i - \bar{x_n} + \bar{x_n})^T \\
&= \sum_{i=1}^{N}((x_i-\bar{x_n})(x_i-\bar{x_n})^T + (x_i-\bar{x_n})\bar{x_n}^T + \bar{x_n}(x_i-\bar{x_n})^T + \bar{x_n}\bar{x_n}^T) \\
&= \sum_{i=1}^{N}((x_i-\bar{x_n})(x_i-\bar{x_n})^T + \sum_{i=1}^{N} x_i\bar{x_n}^T + \sum_{i=1}^{N}\bar{x_n}x_i^T - N\bar{x_n}\bar{x_n}^T
\end{aligned}$$

we know that $\sum_{i=1}^{N} x_i = N\bar{x_n}$, Hence, we can get:

$$\sum_{i=1}^{N} x_i x_i^T = \sum_{i=1}^{N}((x_i-\bar{x_n})(x_i-\bar{x_n})^T + N\bar{x_n}\bar{x_n}^T$$

Therefore:

$$\begin{aligned}
|\sum_{i=1}^{N} x_i x_i^T| &= |\sum_{i=1}^{N}((x_i-\bar{x_n})(x_i-\bar{x_n})^T + N\bar{x_n}\bar{x_n}^T| \\
&= \begin{vmatrix} \frac{1}{N-1}S & \sqrt{N}\bar{x_n}^T \\ -\sqrt{N}\bar{x_n} & I_p \end{vmatrix} = \begin{vmatrix} \frac{1}{N-1}S - N\bar{x_n}^T\bar{x_n}S^{-1}S & \sqrt{N}\bar{x_n}^T \\ 0 & I_p \end{vmatrix} \\
&= |\sum_{i=1}^{N}((x_i-\bar{x_n})(x_i-\bar{x_n})^T|(1+\frac{T^2}{n-1})
\end{aligned}$$

And we can get:

$$LR = (\frac{1}{1 + \frac{T^2}{n-1}})^{\frac{N}{2}}$$

which complete the proof.

(b)

let $Z = \sqrt{n} S_N^{-\frac{1}{2}} \bar{x_n}$. Then we have:

$$T^2 = Z^T Z$$

And, under the null hypothesis, the mean of x is 0.we have:

$$\bar{x_n} \sim N_p(0, \frac{S}{N} \rightarrow Z \sim N_p(0, SS_N^{-1})$$

We know S_N is unbias estimator of S. Hence, when $n \rightarrow \infty$, $S_N \rightarrow S$. And $Z \sim N_p(0, I_p)$ is a standard p-dim normal distribution . Thus:

$$T^2 = Z^T Z = \sum_{i=1}^{p} z_i^2 \sim \chi_p^2$$

which is the end of proof.

(c)

Since we only care about the type I error, all the data will extra from a 3-dim normal distribution with mean 0 and random Σ. We will simulate this process 2000 times to get a estimator of type I error with $alpha = 0.05$.

after running 2000 times with random seed is 42, we get the estimator of type I error is 0.054, which is very close to 0.05. And we think it is a good result and it controls the type I error well.

(d)

With similar setting,we get the result:

1. p=10, Type I error:0.0935
2. p=40, Type I error:0.703
3. p=80, Type I error:1.0

As we can see, with the increse of p, the ability for (b) to control the Type I error become weaker and weaker. For p=40 and 80. It is impossible to say that it controls the type I error well.

(e)

We do the same thing as we did in part(b). However this time the mean of our distribution is not 0. And we will use μ to present it. Now we have:

$$\bar{x_n} \sim N_p(\mu, \frac{S}{N} \rightarrow Z \sim N_p(\sqrt{N} S_N^{-\frac{1}{2}} \mu, SS_N^{-1})$$

As $N \rightarrow \infty$, we get $Z \rightarrow N_p(\sqrt{N} S^{-\frac{1}{2}} N\mu, I_p)$.And, we know $T^2 = \sum_{i=1}^{p} z_i^2$. for every z_i^2, we know z_i^2 is an Noncentral chi-squared distribution with mean to ∞ as $N \rightarrow \infty$ and 1 degree of freedom.

Hence, we can say for every z_i^2, they are iid. And we can rewrite T^2:

$$T^2 = \sum_{i=1}^{p} z_i^2 = p\bar{z_i^2}$$

And we know the mgfs of Noncentral chi-squared distribution with 1 degree of freedom exists in a neighborhood of 0 as well as $p \rightarrow N\gamma, N \rightarrow \infty, N\gamma \rightarrow \infty$, we can apply CLT to $\bar{z_i^2}$. That is to say there exists A_N, B_N, so that:

$$A_N(\bar{z_i^2} - B_N) \sim N(0, 1)$$

Problem 4 continued on next page. . .

Hence:
$$\frac{A_N}{p}(p\bar{z_i^2} - pB_N) \sim N(0,1) \rightarrow a_n(T^2 - b_n) \sim N(0,1)$$

(f)

No. first, if $p > N$, S_n is singular and S_n^{-1} does not exists. And we can also see in part(d), when p becomes close to N, Hotelling's T^2-test become more and more weaker, not to say whan $p > N$. For this part simulation, we will use the method from Muni S. Srivastava [1].We use the similar setting of part(b) with p = 150 and N=100 to get the Type I error is 0.004.

References

[1] Muni S Srivastava and Meng Du. A test for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis*, 99(3):386–402, 2008.