MA554 Applied Multivariate Analysis HW2

Aukkawut Ammartayakun

2023-10-6

Problem 1

Flesh out the proofs of Proposition 1(2) and (6) in Lecture 2 (The Wishart Distribution).

> **Proposition 1 (2)** For $\mathbf{M} \sim W_p(n, \Sigma)$, its characteristic function is $\mathbb{E}\left[\exp(\text{tr}(\mathbf{\Theta}\mathbb{M})i)\right] = |\mathbb{I} - 2i\mathbf{\Theta}\mathbb{M}|^{-\frac{n}{2}}$ for any real, symmetric $p \times p$ matrix $\mathbf{\Theta}$.

Proof We know that

$$\begin{aligned}
\text{tr}(\mathbf{\Theta}\mathbf{M}) &= \text{tr}\left(\mathbf{\Theta}\sum_{j=1}^{n}\mathbf{X}_j\mathbf{X}_j^\top\right) \\
&= \sum_{j=1}^{n}\mathbf{X}_j^\top\mathbf{\Theta}\mathbf{X}_j
\end{aligned}$$

From this, we can see that

$$\begin{aligned}
\mathbb{E}\left[\exp(\text{tr}(\mathbf{\Theta}\mathbb{M})i)\right] &= \mathbb{E}\left[\exp\left(i\sum_{j=1}^{n}\mathbf{X}_j^\top\mathbf{\Theta}\mathbf{X}_j\right)\right] \\
&= \left(\mathbb{E}\left[\exp(i\mathbf{X}_1^\top\mathbf{\Theta}\mathbf{X}_1)\right]\right)^n
\end{aligned}$$

Because $\mathbf{\Theta}^\top = \theta$ and $\Sigma^{-1} > 0, \exists R, \det(R) \neq 0$ and $R^\top\mathbf{\Theta}R = D$ for diagonal matrix D and $R^\top\Sigma R = \mathbb{I}$. Then,

$$\begin{aligned}
\mathbf{X}_1^\top\mathbf{\Theta}\mathbf{X}_1 &= \mathbf{X}_1^\top(R^\top)^{-1}R^\top\mathbf{\Theta}RR^{-1}\mathbf{X}_1 \\
&= (R^{-1}\mathbf{X}_1)^\top D(R^{-1}\mathbf{X}_1^\top)
\end{aligned}$$

Let $Z = R^{-1}X \sim \mathcal{N}_p(\mathbf{0}, \mathbb{I})$, we have

$$\begin{aligned}
\left(\mathbb{E}\left[\exp(i\mathbf{X}_1^\top\mathbf{\Theta}\mathbf{X}_1)\right]\right)^n &= \left(\mathbb{E}\left[\exp(iZ^\top DZ)\right]\right)^n \\
&= \left(\mathbb{E}\left[\exp\left(i\sum_{k=1}^{p}d_k z_k^2\right)\right]\right)^n \\
&= \left(\prod_{k=1}^{p}\mathbb{E}\left[\exp(id_k z_k^2)\right]\right)^n
\end{aligned}$$

Since, $z_k^2 \sim \chi_1^2$, this expression is characteristic function of the chi-squared distribution. That is,

$$\begin{aligned}
\left(\prod_{k=1}^{p} \mathbb{E}\left[\exp(id_k z_k^2)\right]\right)^n &= \left(\prod_{k=1}^{p}(1-2id_j)\right)^{-\frac{n}{2}} \\
&= |\mathbb{I} - 2iD|^{-\frac{n}{2}} \\
&= \left|R^\top \Sigma^{-1} R - 2iR^\top \mathbf{\Theta} R\right|^{-\frac{n}{2}} \\
&= \left(|R|^2 |\Sigma^{-1}|\,|\mathbb{I} - 2i\mathbf{\Theta}\mathbb{M}|\right)^{-\frac{n}{2}} \\
&= |\mathbb{I} - 2i\mathbf{\Theta}\mathbb{M}|^{-\frac{n}{2}}
\end{aligned}$$

Proposition 1 (6) If $\mathbf{M}_1$ and $\mathbf{M}_2$ are independent and satisfy $\mathbf{M}_1 + \mathbf{M}_2 = \mathbf{M} \sim W_p(n, \Sigma)$ and $\mathbf{M}_1 \sim W_p(n_1, \Sigma)$ then $\mathbf{M}_2 \sim W_p(n - n_1, \Sigma)$.

Proof From $\mathbf{M}_1 \perp\!\!\!\perp \mathbf{M}_2$. We can use the property of characteristic functions that

$$\phi_{\mathbf{M}_1 + \mathbf{M}_2}(t) = \phi_{\mathbf{M}_1}(t)\phi_{\mathbf{M}_2}(t)$$

That is

$$\begin{aligned}
|\mathbb{I} - 2i\mathbf{\Theta}\mathbb{M}|^{\frac{n}{2}} &= |\mathbb{I} - 2i\mathbf{\Theta}\mathbb{M}|^{\frac{n_1}{2}} \phi_{\mathbf{M}_2}(t) \\
\phi_{\mathbf{M}_2}(t) &= |\mathbb{I} - 2i\mathbf{\Theta}\mathbb{M}|^{\frac{n-n_1}{2}}
\end{aligned}$$

which is equivalent of saying that $\mathbf{M}_2 \sim W_p(n - n_1, \Sigma)$

Problem 2

Let $\mathbf{X} = (X_1, X_2, X_3)^\top \sim \mathcal{N}_3(\boldsymbol{\mu}, \Sigma)$, where $\boldsymbol{\mu} = (1, -1, 2)^\top$ and

$$\Sigma = \begin{bmatrix} 4 & 0 & -1 \\ 0 & 5 & 0 \\ -1 & 0 & 2 \end{bmatrix}$$

(a) What is the distribution of a random vector $(X_1, X_2)^\top$?

Let $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$. The transformation of $A\mathbf{X}$ is the same as $(X_1, X_2)^\top$ which is $\mathcal{N}_2(A\boldsymbol{\mu}, A\Sigma A^\top)$ which is

```
mu <- c(1, -1, 2)
sigma <- matrix(
        c(4, 0, -1,
          0, 5,  0,
          -1, 0,  2),
        ncol=3, byrow=TRUE
)
A <- matrix(c(1, 0, 0, 0, 1, 0), ncol=3, byrow=TRUE)
A %*% mu
```

```
##      [,1]
## [1,]    1
## [2,]   -1
```

```
A %*% sigma %*% t(A)
```

```
##      [,1] [,2]
## [1,]    4    0
## [2,]    0    5
```

In another word, $(X_1, X_2)^\top \sim \mathcal{N}_2\left(\begin{bmatrix}1 & -1\end{bmatrix}^\top, \begin{bmatrix}4 & 0\\ 0 & 5\end{bmatrix}\right)$

(b) What is the conditional distribution X_1 given $X_3 = x_3$?

We can see that

$$\mathbf{X}_1|\mathbf{X}_3 = x_3 \sim \mathcal{N}(\mu_1 + \Sigma_{13}\Sigma_{33}^{-1}(x_3 - \mu_3), \Sigma_{11} - \Sigma_{13}\Sigma_{33}^{-1}\Sigma_{21})$$

```
new_variance <- sigma[1,1] - sigma[1,3]*1/sigma[3,3]*sigma[2,1]
new_variance
```

```
## [1] 4
```

or $\mathbf{X}_1|\mathbf{X}_3 = x_3 \sim \mathcal{N}(1 - \frac{x_3-2}{2}, 4)$

(c) What is the conditional distribution X_1 given $X_2 = x_2$ and $X_3 = x_3$?

We can use the same method with (b). WLOG, let $Y = (X_2\ X_3)^\top$. We then get

$$\mathbb{E}[X_1|Y = (x_2\ x_3)^\top] = \mu_1 + [\Sigma_{12}\ \Sigma_{13}]\begin{bmatrix}\Sigma_{22} & \Sigma_{23}\\ \Sigma_{32} & \Sigma_{33}\end{bmatrix}^{-1}[x_2 - \mu_2\ x_3 - \mu_3]^\top$$

and

$$\mathrm{Var}[X_1|Y = (x_2\ x_3)^\top] = \Sigma_{11} + [\Sigma_{12}\ \Sigma_{13}]\begin{bmatrix}\Sigma_{22} & \Sigma_{23}\\ \Sigma_{32} & \Sigma_{33}\end{bmatrix}^{-1}[\Sigma_{21}\ \Sigma_{31}]^\top$$

```
variance_three <- sigma[1,1] + c(sigma[1,2],sigma[1,3])%*%
         matrix(c(sigma[2,2], sigma[2,3],sigma[3,2], sigma[3,3]), nrow = 2) %*%
        c(sigma[2,1],sigma[3,1])
variance_three
```

```
##      [,1]
## [1,]    6
```

This $X_1|Y$ is univariate normal distribution with variance of 6 and mean as a function of x_2 and x_3 as noted by expectation above.

Problem 3

Let $\mathbf{X}_1, \ldots \mathbf{X}_n$ be a random sample from $\mathcal{N}_p(\mathbf{0}, \Lambda)$, where Λ is a $p \times p$ diagonal matrix of $\sigma_1^2, \ldots, \sigma_p^2$. Let $D = \mathrm{diag}(s_1^2, \ldots, s_p^2)$ for s_i^2 be the ith diagonal element of the sample variance–covariance matrix. Define a modified Hotelling's statistic $T = \frac{n}{p}\bar{\mathbf{X}}^\top D^{-1}\bar{\mathbf{X}}$ where $\bar{\mathbf{X}} = n^{-1}\sum_{k=1}^n \mathbf{X}_k$

(a) Argue that T is well-defined even when $n < p$

The product can be explicitly written as

$$\begin{aligned} T &= \frac{n}{p}\sum_{i=1}^p \frac{\bar{x}_i^2}{s_i^2} \\ &= \frac{1}{p}\sum_{i=1}^p \left[\frac{\bar{x}_i}{s_i^2}\sum_{j=1}^n x_{ij}\right] \end{aligned}$$

[Intended Skip]

(b) show that the distribution of T can be approximated by a normal distribution, if p is large enough. [Hint: Use central limit theorem for large p.] Given the expression from (a), we can see that it is somewhat the mean of p. From CLT, we can say that distribution of T would converge into asymptotic normal with adjusted mean and variances.

Problem 4

Let $\mathbf{X}_i$ be i.i.d. $\mathcal{N}_2(\boldsymbol{\mu}, \Sigma)$. Given observations $\mathbf{x}_i, i = 1, \ldots, 42$, we have sufficient statistic $\bar{\mathbf{x}} = (.564, .603)^\top$ (sample mean) and $\mathbf{S} = \begin{bmatrix} .0144 & .0117 \\ .0117 & .0146 \end{bmatrix}$ (sample variance). Now consider constructing a confidence region of size $1 - \alpha$ for $\boldsymbol{\mu} = (\mu_1, \mu_2)^\top$.

```
s <- matrix(c(.0144, .0117, .0117, .0146), nrow = 2)
n <- 42
s/n
```

(a) What is MLE of Σ?

```
##              [,1]         [,2]
## [1,] 0.0003428571 0.0002785714
## [2,] 0.0002785714 0.0003476190
```

(b) Evaluate the expression for 95% elliptical confidence region for $\boldsymbol{\mu}$. Denote this region as R_1. Is $\boldsymbol{\mu}_0 = (.60, .58)^\top$ in R_1?

The confidence region would be

```
library(MASS)
x_bar <- matrix(c(.564, .603), nrow = 2)
S <- matrix(c(.0144, .0117, .0117, .0146), nrow = 2)
n <- 42
p <- 2
alpha <- 0.05
F_crit <- ((n-1)*p)/(n-p) * qf(1-alpha, p, n-p)
F_crit*((n-1)*p)/(n-p)
```

```
## [1] 13.58133
```

$$R_1 = \left\{ \boldsymbol{\mu} \in \mathbb{R}^2 | (\bar{\mathbf{X}} - \mu)^\top S^{-1} (\bar{\mathbf{X}} - \mu) < 13.58133 \right\}$$

```
ellipse_val <- function(mut) {
  diff <- mut - x_bar
  t(diff) %*% solve(S) %*% diff
}
mu0 <- matrix(c(.60, .58), nrow = 2)
ellipse_mu0 <- ellipse_val(mu0)
in_ellipse <- ellipse_mu0 <= (n-1)*p/n * F_crit
in_ellipse
```

```
##      [,1]
## [1,] TRUE
```

This conclude that $\boldsymbol{\mu}_0 \in R_1$

```r
CI_mu1 <- x_bar[1] + c(-1,1) * sqrt(S[1,1] * F_crit)
CI_mu2 <- x_bar[2] + c(-1,1) * sqrt(S[2,2] * F_crit)

list(CI_mu1 = CI_mu1, CI_mu2 = CI_mu2)
```

(c) Evaluate the simultaneous confidence intervals for μ_1 and μ_2. Denote this region as R_2. Note that R_2 is a rectangle.

```
## $CI_mu1
## [1] 0.2551302 0.8728698
##
## $CI_mu2
## [1] 0.2919926 0.9140074
```

```r
T2 <- n * t(mu0 - x_bar) %*% solve(S) %*% (mu0 - x_bar)
p_value <- 1-pf(T2, p, n-p)

list(test_statistic = T2, null_distribution = "F(p, n-p)", p_value = p_value)
```

(d) Conduct a hypothesis test for

$$H_0 : \boldsymbol{\mu}_0 = (.60, .58)^\top \;\; vs \; H_1 : \neg H_0$$

Report your (observed) test statistic, the theoretical null distribution, and the p-value.

```
## $test_statistic
##          [,1]
## [1,] 26.29045
##
## $null_distribution
## [1] "F(p, n-p)"
##
## $p_value
##              [,1]
## [1,] 5.137728e-08
```

From the p-value, the test suggest that there is strong statistical evidence against the null hypothesis.

Problem 5

An alternative to the simulataneous confidence interval is the Bonferroni method for multiple comparison. In the previous problem, since there are only two parameters (μ_1, μ_2) involved, we will see that Bonferroni method is advantageous compared to the simultaneous confidence intervals. In general, for m linear combinations $\mathbf{a}_1^\top \boldsymbol{\mu}, \ldots, \mathbf{a}_m^\top \boldsymbol{\mu}$, let C_i denote the confidence interval for the ith linear combination, $\mathbf{a}_i^\top \boldsymbol{\mu}$ with confidence level $1 - \alpha_i$. We have

$$\begin{aligned} P(\mathbf{a}_i^\top \boldsymbol{\mu} \in C_i, \forall i) &= 1 - P(\exists i, \mathbf{a}_i^\top \boldsymbol{\mu} \notin C_i) \\ &\geq 1 - \sum_{i=1}^{m} P(\mathbf{a}_i^\top \boldsymbol{\mu} \notin C_i) \\ &= 1 - (\alpha_1 + \cdots + \alpha_m) \end{aligned}$$

(a) Show that the interval $C_i(\alpha_i) = \mathbf{a}_i^\top \bar{\mathbf{x}} \pm t_{n-1}\left(1 - \frac{\alpha_i}{2}\right)\sqrt{\frac{\mathbf{a}_i^\top \mathbf{S} \mathbf{a}_i}{n}}$ is a confidence interval of $\mathbf{a}_i^\top \boldsymbol{\mu}$ with confidence level $1 - \alpha_i$, where $t_{n-1}\left(1 - \frac{\alpha_i}{2}\right)$ refers to the $100\left(1 - \frac{\alpha_i}{2}\right)\%$ percentile of a t

distribution with $n-1$ degree of freedom.

If $C_i(\alpha_i)$ is a confidence interval, that means,

$$P(\mathbf{a}_i^\top \boldsymbol{\mu} \in C_i) \geq 1 - \alpha_i$$

Now, consider the linear combination $Y_i = \mathbf{a}_i^\top \boldsymbol{\mu}$. It is clear that Y_i is normally distributed. Define $d_i = Y_i - \mathbf{a}_i^\top \bar{\mathbf{x}}$ to be the derivation from the sample mean, we can see that the variance of Y is

$$S_Y^2 = \frac{1}{n-1} \sum_{i=1}^{n} d_i^2$$

Note that S_Y^2 is an estimator of $\mathbf{a}_i^\top \Sigma \mathbf{a}_i$ which is equivalent to $\frac{n}{n-1} \mathbf{a}_i^\top S \mathbf{a}_i$. As shown in the lecture that the ratio

$$Q = \frac{(n-1)S^2}{\mathbf{a}_i^\top \Sigma \mathbf{a}_i}$$

is indeed chi-squared distributed with $n-1$ degree of freedom. Consider the standard error

$$\begin{aligned} SE &= \sqrt{\frac{\mathbf{a}_i^\top \mathbf{S} \mathbf{a}_i}{n}} \\ &= \sqrt{\frac{Q}{n-1}} \end{aligned}$$

By definition of t-distribution, we can see that

$$\frac{\mathbf{a}_i^\top \bar{\mathbf{x}} - \mathbf{a}_i^\top \boldsymbol{\mu}}{SE} \sim t_{n-1}$$

Therefore,

$$P(-t_{n-1,1-\frac{\alpha_i}{2}} \leq \frac{\mathbf{a}_i^\top \bar{\mathbf{x}} - \mathbf{a}_i^\top \boldsymbol{\mu}}{SE} \leq t_{n-1,1-\frac{\alpha_i}{2}}) = 1 - \alpha_i$$

rearrange this and it is clear that this is indeed confidence interval of $\mathbf{a}_i^\top \boldsymbol{\mu}$ with confidence level $1 - \alpha_i$.

(b) Verify the following:

$$P\left(\mathbf{a}_i^\top \boldsymbol{\mu} \in C_i\left(\frac{\alpha}{m}\right), \forall i\right) \geq 1 - \alpha$$

This gives a Bonferroni type simultaneous confidence intervals $C_i\left(\frac{\alpha}{m}\right)$ for level $1-\alpha$.

Since

$$P(\mathbf{a}_i^\top \boldsymbol{\mu} \in C_i, \forall i) \geq 1 - \sum_{i=1}^{m} P(\mathbf{a}_i^\top \boldsymbol{\mu} \notin C_i)$$

Given that

$$\sum_{i=1}^{m} P(\mathbf{a}_i^\top \boldsymbol{\mu} \notin C_i) = \sum_{i=1}^{m} \frac{\alpha}{m} = \alpha$$

Then,

$$P\left(\mathbf{a}_i^\top \boldsymbol{\mu} \in C_i\left(\frac{\alpha}{m}\right), \forall i\right) \geq 1 - \alpha$$

which gives the Bonferroni simultaneous confidence interval.

(c) Using the data in Problem 4, evaluate the simultaneous confidence intervals for μ_1 and μ_2 using the Bonferroni method. Denote this region as R_3.

```
x_bar <- c(.564, .603)
S <- matrix(c(.0144, .0117, .0117, .0146), nrow = 2)
n <- 42
alpha <- 0.05
m <- 2

CI_Bonferroni_mu1 <- x_bar[1] + c(-1, 1) * qt(1-alpha/(2), n-1) * sqrt(S[1,1]/n)
CI_Bonferroni_mu2 <- x_bar[2] + c(-1, 1) * qt(1-alpha/(2), n-1) * sqrt(S[2,2]/n)

list(CI_Bonferroni_mu1 = CI_Bonferroni_mu1, CI_Bonferroni_mu2 = CI_Bonferroni_mu2)
```

```
## $CI_Bonferroni_mu1
## [1] 0.5266054 0.6013946
##
## $CI_Bonferroni_mu2
## [1] 0.5653466 0.6406534
```

(d) Plot the regions R_1, R_2, R_3 in the same figure. Which one do you prefer and why?

```
library(mixtools)
```

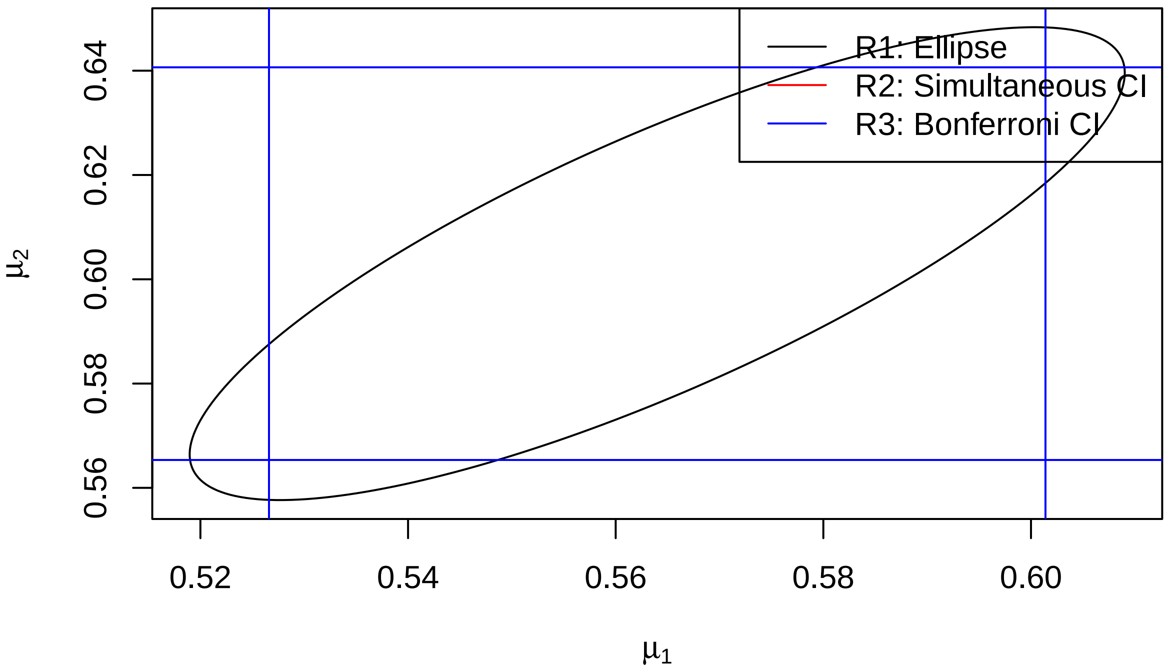
```
## mixtools package, version 2.0.0, Released 2022-12-04
## This package is based upon work supported by the National Science Foundation under Grant No. SES-0518
```

```
p <- 2
q_t <- qt(0.025, n-1, lower.tail = FALSE)

mixtools::ellipse(x_bar, S/n, alpha = (p*(n-1)/(n-p))*qf(0.025,p,n-p),
                  newplot = TRUE, type = 'line', xlab = expression(mu[1]), ylab = expression(mu[2]))
```

```
## Warning in plot.xy(xy, type, ...): plot type 'line' will be truncated to first
## character
```

```
abline(v = CI_mu1, h = CI_mu2, col = "red")
abline(v = CI_Bonferroni_mu1, h = CI_Bonferroni_mu2, col = "blue")
legend("topright", c("R1: Ellipse", "R2: Simultaneous CI", "R3: Bonferroni CI"),
       col = c("black", "red", "blue"), lty = c(1,1,1))
```




```
#not sure why simultaneous not showing up, possibly because it is larger than what is shown.
```

From this, R_3 would be the most preferable due to the range of the interval that is the narrowest.