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

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

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

```
##      [,1]
## [1,]    3
## [2,]    0
```

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

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

In another word, $(X_1, X_2)^\top \sim \mathcal{N}_2\left(\begin{bmatrix} 3 & 0 \end{bmatrix}^\top, \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix}\right)$

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

bla bla

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

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 $D = \text{diag}(s_1^2, \ldots, s_p^2)$ 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 explicilty written as

$$\begin{aligned} T &= \frac{n}{p}\sum_{i=1}^{p}\frac{\bar{x}_i^2}{s_i^2} \\ &= \ldots \end{aligned}$$

(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.]

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 $\boldsymbol{\mu} = (\mu_1, \mu_2)^\top$.

(a) What is MLE of Σ?

(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?

(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.

(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.

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.

(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 $1 - \alpha$.

(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.

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