Multivariate Analysis Lecture 6: Sample Covariance Matrix and Wishart Distribution

Zhaoxia Yu
Professor, Department of Statistics

2023-04-20

Section 1

The Big Picture

The Big Picture: Univariate vs Multivariate

- Review: A random sample, denoted by $X_1, \cdots, X_n$, from a (univariate) normal distribution $N(\mu, \sigma^2)$
  - What are the distributions of $\bar{X}, s^2$? What useful statistics can be constructed?
- New material: A random sample, denoted by $\mathbf{X}_1, \cdots, \mathbf{X}_n$, from a multivariate normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
  - What are the distributions of $\bar{\mathbf{X}}, \mathbf{S}$? What useful statistics can be constructed?

The Big Picture: Univariate

- A random sample, denoted by $X_1, \cdots, X_n$, from a (univariate) normal distribution $N(\mu, \sigma^2)$
- Let $\mathbf{X}_{n\times 1} = (X_1, \cdots, X_n)^T$. It is random vector with a multivarite normal distribution, i.e.,

$$\mathbf{X}_{n\times 1} = (X_1, \cdots, X_n)^T \sim \mathbf{N}(\mu\mathbf{1}, \sigma^2\mathbf{I})$$

1. $\bar{X} \sim N(\mu, \sigma^2/n)$
2. $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$
3. Independence between $\bar{X}$ and s^2.
4. a t-statistic is

$$\frac{\frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}}}{\sqrt{\frac{(n-1)s^2/\sigma^2}{n-1}}} = \frac{\sqrt{n}(\bar{X}-\mu)}{s}$$

It follows the t-distribution with n-1 degrees of freedom, denoted by t_{n-1}.

The Big Picture: Multivariate

- A random sample $\mathbf{X}_1, \cdots, \mathbf{X}_n$ from a multivariate normal distribution $\mathbf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
- Let

$$\mathbf{X}_{n\times p} = \begin{pmatrix} \mathbf{X}_1^T \\ \vdots \\ \mathbf{X}_n^T \end{pmatrix}$$

  $\mathbf{X}$ follows a matrix normal distribution.

1. Sample mean vector follows a multivariate normal, i.e., $\bar{\mathbf{X}} \sim \mathbf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}/n)$
2. Sample covariance matrix $(n-1)\mathbf{S}$ follows a Wishart distribution, i.e., $(n-1)\mathbf{S} \sim Wishart_p(n-1, \Sigma)$
3. Independence between $\bar{\mathbf{X}}$ and S.
4. Hoetelling's T^2: $T^2 = (\bar{\mathbf{X}} - \boldsymbol{\mu})^T \left(\frac{\mathbf{S}}{n}\right)^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu})$

The Big Picture: outline

- Sample variance and chi-squared distribution
- Sample covariance matrix and Wishart distribution
- Hotelling's T^2
- Maximum likelihood estimate

Section 2

Sample Variance

Sample Variance and Chi-squared Distribution

- Let $\mathbf{X} = (X_1, \cdots, X_n)$ denote a random sample from $N(\mu, \sigma^2)$.
- Equivalently, $\mathbf{X} \sim N(\mu\mathbf{1}, \sigma^2\mathbf{I})$.
- Let $s^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ denote the sample variance.
- We would like to show that

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

- Outline of proof
  1. Projection matrices
  2. Chi-squared distribution
  3. Rewrite $(n-1)s^2/\sigma^2$ as the sum of squared $N(0,1)$ random variables

Projection Matrices

- A projection matrix is a square matrix that is both idempotent and symmetric

$$\mathbf{P}^2 = \mathbf{P},\ \mathbf{P} = \mathbf{P}^T$$

Projection Matrices

- Suppose $\mathbf{P}$ is a projection matrix. We have
  - The eigenvalues of $\mathbf{P}$ has eigenvalues are either 0 or 1, and the number of 1's is the same as the rank of the projection matrix.
  - $tr(\mathbf{P}) = rank(\mathbf{P})$
  - The spectral decomposition of $\mathbf{P}$ is

$$\mathbf{P} = \sum_{i=j}^{r} \gamma_j \gamma_j^T$$

where $r = rank(\mathbf{P})$, and $(\gamma_1, \cdots, \gamma_r)$ are orthogonal vectors of norm 1, i.e.,

$$\gamma_i^T \gamma_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

A Special Projection Matrix: the Centering Matrix

- The centering matrix $\mathbb{C} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T$ is a very special matrix.
- It is a projection matrix, which is defined as both symmetric and idempotent:
  - $\mathbb{C}^T = \mathbb{C}$ (symmetric)
  - $\mathbb{C}^2 = \mathbb{C}$ (idempotent)
- One important result about a projection matrix is that its eigenvalues are either zero or one.
- By properties of projection matrices, we have
  - $rank(\mathbb{C}) = tr(\mathbb{C}) = n - 1$
  - $\mathbb{C} = \sum_{j=1}^{n-1} \gamma_j \gamma_j^T$

A Special Projection Matrix: the Centering Matrix

- The centering matrix centers data
- Univariate: Let $\mathbf{X}_{n\times 1}$ be a random sample from $N(\mu, \sigma^2)$, i.e.,

$$\mathbf{X}_{n\times 1} \sim N(\mu\mathbf{1}, \sigma^2\mathbf{I})$$

$\mathbb{C}\mathbf{X}$ is a linear function of $\mathbf{X}$ and it can be verified that $\mathbb{C}\mathbf{1} = \mathbf{0}$, we have

$$E[\mathbb{C}\mathbf{X}] = \mu\mathbb{C}\mathbf{1} = \mathbf{0}$$

- Multivariate: Let $\mathbf{X}_{n\times p}$ be a random sample from $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ Similarly, it can be shown that $\mathbb{C}\mathbf{X}$ has mean $\mathbf{0}_{n\times p}$. We have verified this numerically.
- In either situation, we have $\mathbb{C}\mathbf{X} = \mathbb{C}(\mathbf{X} - E[\mathbf{X}])$ This fact will be used later.

Chi-squared distribution

- Definition. Let $Z_1, Z_2, ..., Z_k$ be independent standard normal random variables. Then, the sum of squares $Q = Z_1^2 + Z_2^2 + ... + Z_k^2$ has a chi-squared distribution with k degrees of freedom, denoted by χ_k^2.
- Alternative definition. Let $\mathbf{Z}_{k\times 1} \sim N(\mathbf{0}, \mathbf{I})$. We say $||\mathbf{Z}||^2 = \mathbf{Z}^T\mathbf{Z}$ follows χ_k^2.
- The PDF of a chi-squared random variable with k degrees of freedom is given by:

$$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2},\ x > 0$$

where $\Gamma(\cdot)$ is the gamma function.

Chi-squared distribution

- The chi-squared distribution is a special case of the gamma distribution, where the shape parameter is $k/2$ and the rate parameter is $1/2$.
- The MGF of a chi-squared random variable with k degrees of freedom is:
$$M_X(t) = (1-2t)^{-k/2}$$
- The mean and variance of a chi-squared random variable with k degrees of freedom are:

$$\mathrm{E}[X] = k, \ \mathrm{Var}[X] = 2k$$

Construct Chi-squared R.V.s using Normal R.V.s and Projection Matrices

- Let $\mathbf{P}_{n\times n}$ be a projection matrix with rank r and let $\mathbf{Z}_{n\times 1} \sim N(\mathbf{0}, \mathbf{I})$

$$\mathbf{Z}^T\mathbf{P}\mathbf{Z} = \mathbf{Z}^T\sum_{i=1}^{r}\gamma_i\gamma_i^T\mathbf{Z} = \sum_{i=1}^{r}\mathbf{Z}^T\gamma_i\gamma_i^T\mathbf{Z}$$
$$= \sum_{i=1}^{r}(\gamma_i^T\mathbf{Z})^T(\gamma_i^T\mathbf{Z})$$

Let $Y_i = \gamma_i^T\mathbf{Z}$. Note that Y_i is univariate and it is a linear combination of $\mathbf{Z}$, from which we can show that $Y_i \sim N(0,1)$.

- Note that $\mathbf{Z}^T\mathbf{P}\mathbf{Z} = \sum_{i=1}^{r} Y_i^2$. By the definition of chi-squared distribution, we have $\mathbf{Z}^T\mathbf{P}\mathbf{Z} \sim \chi^2_r$

The Sample Variance

- We have shown that
  - $\mathbb{C} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T$
  - $\mathbb{C}^T = \mathbb{C}$, $\mathbb{C}^2 = \mathbb{C}$.
  - It is a projection matrix with rank $n-1$ and

$$\mathbb{C} = \sum_{j=1}^{n-1} \gamma_i \gamma_i^T$$

-The he centering matrix does center data, i.e.,

$$\mathbb{C}\mathbf{X} = \mathbb{C}(\mathbf{X} - E[\mathbf{X}])$$

  - $(n-1)s^2 = \mathbf{X}^T \mathbb{C}\mathbf{X}$, where

The Sample Variance

- Therefore,

$$\begin{aligned}\frac{(n-1)s^2}{\sigma^2} &= \frac{(\mathbf{X}-E[\mathbf{X}])^T}{\sigma}\mathbb{C}^T\mathbb{C}\mathbb{C}\frac{(\mathbf{X}-E[\mathbf{X}])}{\sigma}\\ &= \frac{(\mathbf{X}-E[\mathbf{X}])^T}{\sigma}\mathbb{C}\frac{(\mathbf{X}-E[\mathbf{X}])}{\sigma}\end{aligned}$$

The Sample Variance

- Let

$$\mathbf{Z} = \frac{(\mathbf{X} - E[\mathbf{X}])}{\sigma}$$

- Easy to see that $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I})$. Thus,

$$\frac{(n-1)s^2}{\sigma^2} = \mathbf{Z}^T \mathbb{C} \mathbf{Z}$$

- Use the result in previous slides, we have

$$\frac{(n-1)s^2}{\sigma^2} = \mathbf{Z}^T \mathbb{C} \mathbf{Z} \sim \chi^2_{n-1}$$

Section 3

Sample Covariance

The Sample Covriance from A MVN Random Sample

- Let $\mathbf{X}_1, \cdots, \mathbf{X}_n \overset{iid}{\sim} N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
- Recall that the sample covariance matrix is defined as

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

- We have shown that

$$(n-1)\mathbf{S} = \mathbf{X}^T \mathbb{C} \mathbf{X}$$

where $\mathbf{X}$ is the $n \times p$ random matrix.

The Sample Covriance from A MVN Random Sample

- The goal is to show that $(n-1)\mathbf{S}$ follows a Wishart distribution. More specifically, we would like to show that

$$(n-1)\mathbf{S} \sim Wishart_p(n-1, \Sigma)$$

- Outline of proof
  1. Wishart-distribution
  2. Rewrite $(n-1)\mathbf{S}$
  3. Apply properties of a projection matrix
  4. Use the definition of Wishart distribution

Wishart Distribution

- The Wishart distribution is named after the British statistician John Wishart, who introduced it in his 1928 paper published in Biometrika.
- Wishart was interested in the problem of estimating the covariance matrix of a multivariate normal distribution.
- Wishart showed that the sample covariance matrix follows a particular probability distribution that we now call the Wishart distribution.
- The Wishart distribution has become a fundamental tool in multivariate statistical analysis

Definition of Wishart Distribution

- A Wishart distribution can be defined in the following way
- Let $\mathbf{W}$ be a $p \times p$ random matrix. We say $\mathbf{W}$ follows $Wishart_p(k, \mathbf{\Sigma})$ if $\mathbf{W}$ can be written as $\mathbf{W} = \mathbf{X}^T\mathbf{X}$ where $\mathbf{X}$ denotes the random matrix formed by a random sample of size k from MVN $N(\mathbf{0}, \mathbf{\Sigma})$.
- The definition indicates that if we have a random sample $\mathbf{X}_1, \cdots \mathbf{X}_k$ from $N(\mathbf{0}, \mathbf{\Sigma})$, then $\mathbf{X}^T\mathbf{X} = \sum_{i=1}^k \mathbf{X}_i\mathbf{X}_i^T \sim Wishart_p(k, \mathbf{\Sigma})$.
- Remark:$E[\mathbf{W}] = k\Sigma$.

Wishart vs Chi-squared

- Wishart: If $\mathbf{X}_1, \cdots \mathbf{X}_k \overset{iid}{\sim} N(\mathbf{0}, \mathbf{\Sigma})$, then

$$\mathbf{X}^T\mathbf{X} = \sum_{i=1}^{k} \mathbf{X}_i\mathbf{X}_i^T \sim Wishart_p(k, \mathbf{\Sigma}), \text{ where } \mathbf{X}_{k\times p} = \begin{pmatrix} X_1^T \\ \vdots \\ X_k^T \end{pmatrix}$$

- Chi-squared: If $X_1, \cdots, X_k \overset{iid}{\sim} N(0, 1)$, then

$$\mathbf{X}^T\mathbf{X} = \sum_{i=1}^{k} X_i^2 \sim \chi_k^2, \text{ where } \mathbf{X}_{k\times 1} = \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix}$$

Wishart vs Chi-squared (continued)

- When $p = 1$,

$$W = \sum_{i=1}^{k} X_i^2 = \sigma^2 \sum_{i=1}^{k} \left(\frac{X_i}{\sigma} \right)^2 \sim \sigma^2 \chi_k^2$$

The Sample Covariance Matrix

- Let $\mathbf{X}_1, \cdots \mathbf{X}_n$ be a random sample from $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The $\mathbf{X}_{n \times p}$ follows a matrix normal distribution:

$$\mathbf{X} \sim N(\mathbf{1}_n \otimes \boldsymbol{\mu}^T, \boldsymbol{\Sigma}, \mathbf{I}_n)$$

- The sample covariance $(n-1)\mathbf{S} = \mathbf{X}^T \mathbb{C} \mathbf{X}$ is based on the centered data. The definition of Wishart distribution is not applicable immediately.
- Next we show that $(n-1)\mathbf{S}$ follows $Wishart_p(n-1, \boldsymbol{\Sigma})$.

The Sample Covariance Matrix

- Rewrite $(n-1)\mathbf{S}$:

$$
\begin{aligned}
(n-1)\mathbf{S} &= \mathbf{X}^T\mathbb{C}^T\mathbb{C}\mathbb{C}\mathbf{X} = (\mathbb{C}\mathbf{X})^T(\mathbb{C}\mathbf{X}) \\
&= (\mathbb{C}\mathbf{X})^T\mathbb{C}(\mathbb{C}\mathbf{X}) \\
&= (\mathbb{C}\mathbf{X})^T\sum_{j=1}^{n-1}\gamma_i\gamma_i^T(\mathbb{C}\mathbf{X}) \\
&= \sum_{j=1}^{n-1}(\gamma_i^T\mathbb{C}\mathbf{X})^T(\gamma_i^T\mathbb{C}\mathbf{X})
\end{aligned}
$$

The Sample Covariance Matrix

- Let $Y_i = (\gamma_i^T \mathbb{C}\mathbf{X})^T$, we have
  - $E[Y_i] = 0$ because $\mathbb{C}$ is the centering matrix
  - In the following, we show that Y_i and Y_j are uncorrelated for $i \neq j$:

$$
\begin{aligned}
Cov[Y_i, Y_j] &= E[(Y_i - \mathbf{0})(Y_j - \mathbf{0})^T] \\
&= E[Y_i Y_j^T] \\
&= E[(\gamma_i^T \mathbb{C}\mathbf{X})^T (\gamma_j^T \mathbb{C}\mathbf{X})] \\
&= E[\mathbf{X}^T \mathbb{C}\gamma_i \gamma_j^T \mathbb{C}\mathbf{X}] \\
&= \mathbf{0}
\end{aligned}
$$

The last step is true because for $i \neq j$, $\gamma_i \gamma_j^T = 0$

The Sample Covariance Matrix

Since Y_i and Y_j are two linear combinations of the same MVN distributed random matrix (or its vectorized version), we have Y_i and Y_j are independent for $i \neq j$.

- It can also be shown that $Y_i \sim N(\mathbf{0}, \Sigma)$.
- By the definition of Wishart, we can conclude that

$$(n-1)\mathbf{S} \sim Wishart_p(n-1, \boldsymbol{\Sigma})$$

Some Interesting Results

- Consider a random sample from MVN $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let $\mathbf{S}$ denote the sample covariance matrix.
- We have already shown that $(n-1)\mathbf{S} \sim Wishart_p(n-1, \boldsymbol{\Sigma})$
- What is the distribution of a diagonal element of $(n-1)\mathbf{S}$?
- What is the distribution of the sum of elements of $(n-1)\mathbf{S}$?
- What is the distribution of $(n-1)\mathbf{BSB}^T$ where B is a fixed $q \times p$ matrix?
- If time permits, we will run some simulations

Some Interesting Results (continued)

- If you cannot get the answer to the last question, let's use the definition of Wishart distribution.
- Let $\mathbf{W} = (n-1)S$. Because it follows $Wishart_p(n-1, \mathbf{\Sigma})$, we know that $\mathbf{W} = \sum_{j=1}^{n-1} \mathbf{Z}_j \mathbf{Z}_j^T$ where $\mathbf{Z}_j$'s are iid frm $N(\mathbf{0}, \mathbf{\Sigma})$.
- Then

$$
\begin{aligned}
(n-1)\mathbf{BSB}^T = \mathbf{B} \sum_{j=1}^{n-1} \mathbf{Z}_j \mathbf{Z}_j^T \mathbf{B} &= \sum_{j=1}^{n-1} \mathbf{B} \mathbf{Z}_j \mathbf{Z}_j^T \mathbf{B}^T \\
&= \sum_{j=1}^{n-1} (\mathbf{B}\mathbf{Z}_j)(\mathbf{B}\mathbf{Z}_j)^T
\end{aligned}
$$

Some Interesting Results (continued)

Let $\mathbf{Y}_j = \mathbf{B}\mathbf{Z}_j$. Note that it is a linear function of $\mathbf{Z}_j$; therefore

$$\mathbf{Y}_j \sim N(\mathbf{0}, \mathbf{B}\mathbf{\Sigma}\mathbf{B}^T)$$

and the $\mathbf{Y}_j$'s are iid (becaue ...).

By the definition of Wishart distribution, we have

$$(n-1)\mathbf{B}\mathbf{S}\mathbf{B}^T \sim Wishart_q(n-1, \mathbf{B}\mathbf{\Sigma}\mathbf{B}^T)$$

Section 4

Hotelling's T^2

The Hotelling's T^2 Statistic

- Finally we are ready to introduce Hotelling's
- The student's t is used for making inference of mean(s) of normal distribution(s)
- Hotelling generalized the student's t, which is for univarite, to Hotelling's T2, which the multivariate version

Definition Hotelling's T^2

- Definition. We say a random variable follows Hotelling's T^2 $T^2_{p,\nu}$ if the random variable can be written as $\mathbf{Z}^T \left(\frac{W}{\nu}\right)^{-1} \mathbf{Z}$ where
  1. $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{\Sigma})$
  2. $\mathbf{W} \sim W_p(\nu, \boldsymbol{\sigma})$
  3. $\mathbf{Z} \perp \mathbf{W}$

One-Sample Hotelling T^2

- Let $\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_n$ be a random sample from a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.
- The sample mean vector and sample covariance matrix are denoted by $\bar{\mathbf{X}}$ and $\mathbf{S}$, respectively.
- The null hypothesis of interest $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$
- The one-sample Hotelling T^2 is defined as

$$T^2 = (\hat{\mu} - \mu_0)^T \left(Cov(\hat{\mu})\right)^{-1} (\hat{\mu} - \mu_0)$$

One-Sample Hotelling T^2 (continued)

- To see that T^2 does follow Hotelling's T^2, we rewrite it

$$\begin{aligned}
T^2 &= (\hat{\mu} - \mu_0)^T (Cov(\hat{\mu}))^{-1} (\hat{\mu} - \mu_0) \\
&= (\bar{\mathbf{X}} - \mu_0)^T \left(Cov(\bar{\mathbf{X}})\right)^{-1} (\bar{\mathbf{X}} - \mu_0) \\
&= (\bar{\mathbf{X}} - \mu_0)^T \left(\frac{S}{n}\right)^{-1} (\bar{\mathbf{X}} - \mu_0) \\
&= [\sqrt{n}(\bar{\mathbf{X}} - \mu_0)]^T \left(\frac{(n-1)S}{n-1}\right)^{-1} [\sqrt{n}(\bar{\mathbf{X}} - \mu_0)]
\end{aligned}$$

- We have shown that all the three conditions for constructing a Hotelling's T^2 are satisfied
- As a result, $T^2 \sim T^2_{p,n-1}$.

Hotelling's T^2 Distribution vs F Distribution

Hotelling's T^2

Claim: $T^2_{p,\nu} \sim \frac{\nu p}{\nu+1-p} F_{p,\nu+1-p}$.

For the T^2 statistic, we have $T^2 \overset{H_0}{\sim} \frac{(n-1)p}{n-p} F_{p,n-p}$. We reject H_0 at significance level α when $T^2 > \frac{(n-1)p}{n-p} F_{p,n-p,1-\alpha}$.

Corollary.

$$\frac{n-p}{p}(\bar{X}-\mu_0)^T(\hat{\Sigma})^{-1}(\bar{X}-\mu_0) \overset{H_0}{\sim} F_{p,n-p}$$

where $\hat{\Sigma} = \frac{1}{n}X^T H X = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T = \frac{(n-1)S}{n}$.

Section 5

MLE

MLE: Introduction

- The maximum likelihood estimate (MLE) is a widely used method for estimating the parameters of a statistical model.
- In this presentation, we will focus on the MLE for a multivariate normal distribution.

MLE: Multivariate Normal Distribution

- A random vector $\boldsymbol{X}$ follows a p-dimensional multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, denoted by $\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, if its probability density function is given by:

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

where $|\boldsymbol{\Sigma}|$ denotes the determinant of $\boldsymbol{\Sigma}$.

MLE:Maximum Likelihood Estimate

- Let $\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_n$ be a random sample from a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.
- The log-likelihood function for the sample is given by:

$$\ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{X}_i - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_i - \boldsymbol{\mu})$$

- The MLE of $\boldsymbol{\mu}$ is the sample mean $\bar{\boldsymbol{X}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i$.

MLE:Maximum Likelihood Estimate (continued)

- To derive the MLE of $\mathbf{\Sigma}$, we first take the derivative of the log-likelihood function with respect to $\mathbf{\Sigma}$ and set it equal to zero:

$$\frac{\partial \ell}{\partial \mathbf{\Sigma}} = -\frac{n}{2}\mathbf{\Sigma}^{-1} + \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{X}_i - \boldsymbol{\mu})(\boldsymbol{X}_i - \boldsymbol{\mu})^T\mathbf{\Sigma}^{-2} = 0$$

- Solving for $\mathbf{\Sigma}$, we obtain the MLE as:

$$\hat{\mathbf{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}})(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}})^T$$

- where $\hat{\boldsymbol{\mu}}$ is the MLE of $\boldsymbol{\mu}$, as previously derived.