Multivariate Analysis
Chapter 4: Multivariate Normal Distribution

Shaobo Jin

Department of Mathematics

Intended Learning Outcome

Through this chapter, you should be able to

1. describe the definition of multivariate normal,
2. memorize the density function of multivariate normal,
3. apply properties of multivariate normal,
4. be aware of Wishart distribution,
5. apply properties of Wishart distribution.

Univariate Normal Distribution

The univariate normal distribution with mean μ and variance σ^2 has the probability density function

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad -\infty < x < \infty.$$

- We often denote it by $X \sim N\left(\mu, \sigma^2\right)$.
- When $X \sim N(0,1)$, we say standard normal distribution.
- A ▸ Galton board "yields" the shape of a univariate normal density function.

From Univariate to Multivariate Normal

Let $Z \sim N(0,1)$. Then, $X = aZ + \mu \sim N\left(\mu, \sigma^2\right)$.

Let $\boldsymbol{Z} = \begin{bmatrix} Z_1 & Z_2 & \cdots & Z_p \end{bmatrix}^T$ be a random vector, each $Z_j \sim N(0,1)$, and Z_j is independent of Z_k for any $j \neq k$. Let $\boldsymbol{A}$ be a constant matrix and $\boldsymbol{\mu}$ be a constant vector. Then,

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{Z} + \boldsymbol{\mu}$$

follows a $p-$dimensional multivariate normal distribution. It is denoted by $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

From Univariate to Multivariate Normal: Density

The density function of the random variable $X \sim N\left(\mu, \sigma^2\right)$ with $\sigma > 0$ can be expressed as

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\left(x-\mu\right)^2}{2\sigma^2}\right\} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(x-\mu\right)\frac{1}{\sigma^2}\left(x-\mu\right)\right\}.$$

A p-dimensional random variable $\boldsymbol{X} \sim N_p\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$ with $\boldsymbol{\Sigma} > 0$ has the density

$$f\left(\boldsymbol{x}\right) = \frac{1}{\left(2\pi\right)^{p/2}\sqrt{\det\left(\boldsymbol{\Sigma}\right)}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\right\},$$

where $-\infty < x_j < \infty$, $j = 1, 2, ..., p$. Here $\mathbb{E}\left(\boldsymbol{X}\right) = \boldsymbol{\mu}$ and $\operatorname{cov}\left(\boldsymbol{X}\right) = \boldsymbol{\Sigma}$.

Special Case: Bivariate Normal

For $p = 2$, let

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} > 0.$$

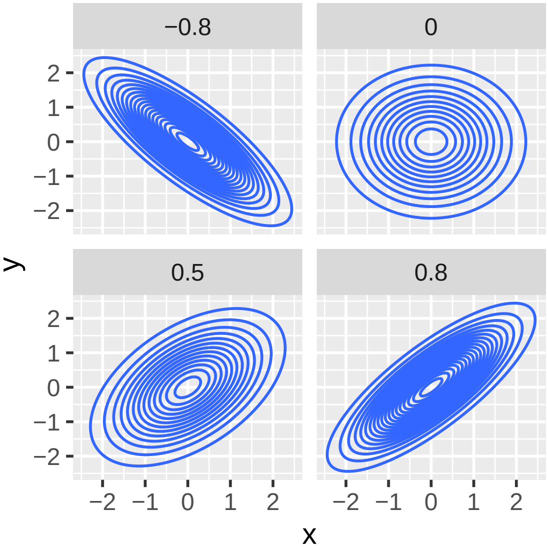
The density of a bivariate normal random vector is

$$\begin{aligned} f(\boldsymbol{x}) &= \frac{1}{(2\pi)^{p/2}\sqrt{\det(\boldsymbol{\Sigma})}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\} \\ &= \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}\left(1-\rho_{12}^2\right)}} \exp\left\{-\frac{1}{2}u^2\right\}, \end{aligned}$$

where

$$u^2 = \frac{1}{1-\rho_{12}^2}\left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 - 2\rho_{12}\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sqrt{\sigma_{11}\sigma_{22}}} + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2\right].$$

Contour of Bivariate Normal Density



Linear Combination(s)

Result 4.2: Linear Combination of Normal is Normal

1. If $\boldsymbol{X} \sim N_p\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$, then $\boldsymbol{a}^T \boldsymbol{X} \sim N\left(\boldsymbol{a}^T \boldsymbol{\mu}, \boldsymbol{a}^T \boldsymbol{\Sigma} \boldsymbol{a}\right)$ for every $\boldsymbol{a}$.
2. If $\boldsymbol{a}^T \boldsymbol{X} \sim N\left(\boldsymbol{a}^T \boldsymbol{\mu}, \boldsymbol{a}^T \boldsymbol{\Sigma} \boldsymbol{a}\right)$ for every $\boldsymbol{a}$, then $\boldsymbol{X}$ must be $N_p\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$.
3. If $\boldsymbol{X} \sim N_p\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$, $\boldsymbol{A}$ is a $q \times p$ matrix of constants, and $\boldsymbol{d}$ is a $p \times 1$ vector of constants, then $\boldsymbol{A}\boldsymbol{X} + \boldsymbol{d} \sim N_q\left(\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{d}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T\right)$.

Corollary: Marginal Distribution is Normal

The marginal distribution of X_i is $N\left(\mu_i, \sigma_{ii}\right)$.

Example

Linear Combinations

Suppose that

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}\right).$$

Find the joint distribution of $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$.

Normal And Chi-Square

If univariate $X \sim N\left(\mu, \sigma^2\right)$, then

$$\frac{(X-\mu)^2}{\sigma^2} \quad \sim \quad \chi_1^2.$$

Result 4.7

Let $\boldsymbol{X}$ be distributed as $N_p\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$ with $\det\left(\boldsymbol{\Sigma}\right) > 0$. Then

$$\left(\boldsymbol{X}-\boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{X}-\boldsymbol{\mu}\right) \quad \sim \quad \chi_p^2,$$

the chi-square distribution with p degrees of freedom.

The quadratic form $\left(\boldsymbol{x}-\boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)$ creates an ellipsoid and $\left(\boldsymbol{x}-\boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right) \leq \chi_p^2\left(\alpha\right)$ occurs with probability $1-\alpha$.

Subset of Variables

Result 4.4

Suppose that $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then, all subsets of $\boldsymbol{X}$ are normally distributed. If we respectively partition $\boldsymbol{X}$, its mean vector $\boldsymbol{\mu}$, and its covariance matrix $\boldsymbol{\Sigma}$ as

$$
\begin{aligned}
\boldsymbol{X}_{p\times 1} &= \begin{bmatrix} \boldsymbol{X}_1\,(q\times 1) \\ \boldsymbol{X}_2\,((p-q)\times 1) \end{bmatrix} \\
\boldsymbol{\mu}_{p\times 1} &= \begin{bmatrix} \boldsymbol{\mu}_1\,(q\times 1) \\ \boldsymbol{\mu}_2\,((p-q)\times 1) \end{bmatrix} \\
\boldsymbol{\Sigma} &= \begin{bmatrix} \boldsymbol{\Sigma}_{11}\,(q\times q) & \boldsymbol{\Sigma}_{12}\,(q\times (p-q)) \\ \boldsymbol{\Sigma}_{21}\,((p-q)\times q) & \boldsymbol{\Sigma}_{22}\,((p-q)\times (p-q)) \end{bmatrix}
\end{aligned}
$$

then $\boldsymbol{X}_1$ is distributed as $N_q(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$, and $\boldsymbol{X}_2$ is distributed as $N_{p-q}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$.

Example: Subset of Variables

Marginal Distribution

Suppose that

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \sim N_2\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} \right).$$

1. Find the distribution of $\begin{bmatrix} X_1 \\ X_3 \end{bmatrix}$.
2. Find the distribution of

$$\begin{bmatrix} X_1 & X_3 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{13} \\ \sigma_{31} & \sigma_{33} \end{bmatrix}^{-1} \begin{bmatrix} X_1 \\ X_3 \end{bmatrix}.$$

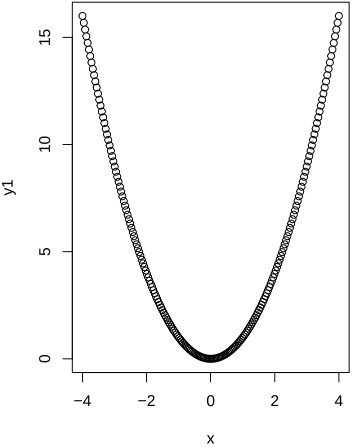
Subset of Variables

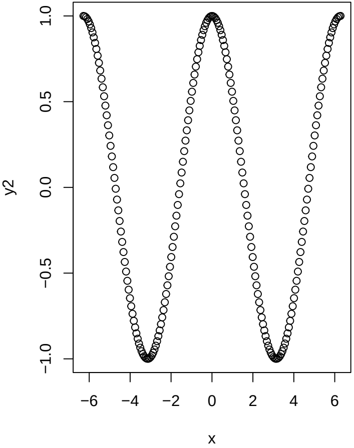
Result 4.5

1. If $\boldsymbol{X}_1$ $(q_1 \times 1)$ and $\boldsymbol{X}_2$ $(q_2 \times 1)$ are independent random vectors (regardless of their distributions), then $\operatorname{cov}(\boldsymbol{X}_1, \boldsymbol{X}_2) = \mathbf{0}$, a $q_1 \times q_2$ matrix of zeros.
2. If $\begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix}$ is a $N_{q_1+q_2}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}\right)$, then $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent if and only if $\boldsymbol{\Sigma}_{12} = \mathbf{0}$.
3. If $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent and are distributed $N_{q_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ and $N_{q_2}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$, respectively, then $\begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix}$ has the multivariate normal distribution
$$N_{q_1+q_2}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}\right).$$

Counterexample

Zero covariance does not necessarily mean independence.





Marginal Normal and Joint Distribution

Marginal normal does not mean that they are joint normal.

Marginal Normal

Suppose that (X, Y) has the joint density function

$$f(x, y) = \frac{1}{2\pi} \exp\left\{-\frac{x^2 + y^2}{2}\right\} [1 + \sin(x) \sin(y)],$$

where $-\infty < x, y < \infty$. Find the marginal distributions.

Conditional Distribution

Result 4.6

Let $\begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix}$ be distributed as $N_p\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}\right)$. Then the conditional distribution of $\boldsymbol{X}_1$ given that $\boldsymbol{X}_2 = \boldsymbol{x}_2$, is

$$\boldsymbol{X}_1 \mid \boldsymbol{X}_2 = \boldsymbol{x}_2 \quad \sim \quad N\left\{\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\left(\boldsymbol{x}_2 - \boldsymbol{\mu}_2\right), \, \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right\},$$

provided that $\boldsymbol{\Sigma}_{22}$ is invertible. Note that the conditional covariance does not depend on $\boldsymbol{x}_2$.

Bivariate normal

Let $\boldsymbol{X}$ be a bivariate normal random vector. Then,

$$X_1 \mid X_2 = x_2 \quad \sim \quad N\left(\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}\left(x_2 - \mu_2\right), \, \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}\right).$$

Conditional Distribution: Example

All Conditional Distribution Still Normal?

Suppose that $\boldsymbol{X} \sim N_p\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$. Let $\boldsymbol{X}_1$ be a subset of $\boldsymbol{X}$, not necessarily contiguous. Let $\boldsymbol{X}_{-1}$ be the supplement. Is the conditional distribution of $\boldsymbol{X}_1$ given $\boldsymbol{X}_{-1}$ normal?

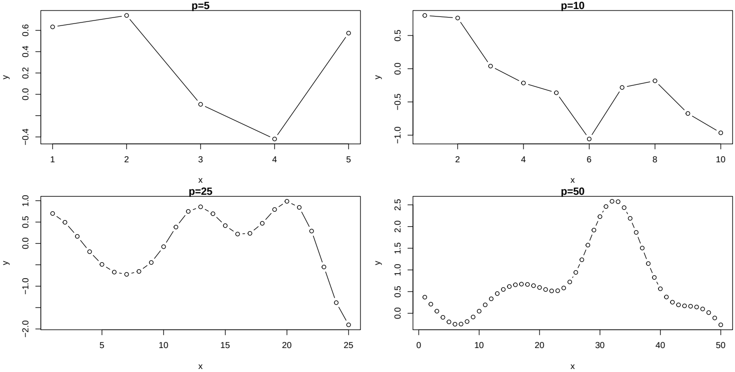
Prediction Using Conditional Mean

Suppose that

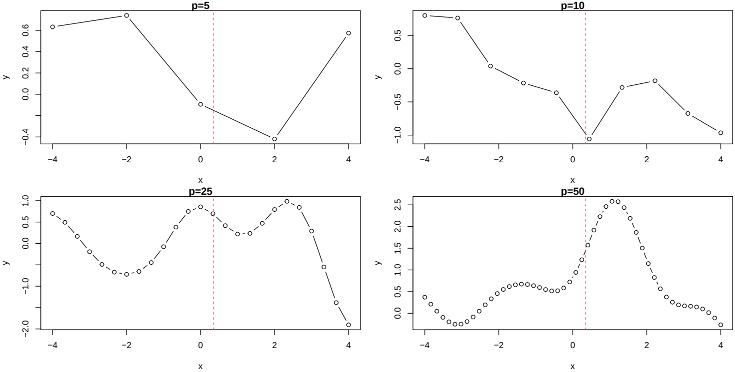
$$\begin{bmatrix} Y_1 \\ \boldsymbol{Y}_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \sigma^2 & \boldsymbol{\rho} \\ \boldsymbol{\rho}^T & \boldsymbol{\Sigma} \end{bmatrix}\right).$$

Find the conditional distribution of Y_1 given $\boldsymbol{Y}_2$.

Multivariate Normal As a Curve



Gaussian Process



Independence and Linear Combination

Result 4.8

Let $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, ..., $\boldsymbol{X}_n$ be mutually independent with $\boldsymbol{X}_j$ distributed as $N_p\left(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}\right)$. Then

$$\boldsymbol{V}_1 = \sum_{j=1}^{n} c_j \boldsymbol{X}_j \sim N_p\left(\sum_{j=1}^{n} c_j \boldsymbol{\mu}_j, \sum_{j=1}^{n} c_j^2 \boldsymbol{\Sigma}\right).$$

Moreover, $\boldsymbol{V}_1$ and $\boldsymbol{V}_2 = b_1 \boldsymbol{X}_1 + b_2 \boldsymbol{X}_2 + \cdots + b_n \boldsymbol{X}_n$ satisfy

$$\begin{bmatrix} \boldsymbol{V}_1 \\ \boldsymbol{V}_2 \end{bmatrix} \sim N_{2p}\left(\begin{bmatrix} \sum_{j=1}^{n} c_j \boldsymbol{\mu}_j \\ \sum_{j=1}^{n} b_j \boldsymbol{\mu}_j \end{bmatrix}, \begin{bmatrix} \sum_{j=1}^{n} c_j^2 \boldsymbol{\Sigma} & \sum_{j=1}^{n} c_j b_j \boldsymbol{\Sigma} \\ \sum_{j=1}^{n} c_j b_j \boldsymbol{\Sigma} & \sum_{j=1}^{n} b_j^2 \boldsymbol{\Sigma} \end{bmatrix}\right).$$

Consequently, $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ are independent if $\sum_{j=1}^{n} c_j b_j = 0$.

Summary

Let a random vector $\boldsymbol{X}_{p\times 1}$ follow a multivariate normal distribution.

- Result 4.2 and Result 4.3: Linear combinations of the components of $\boldsymbol{X}$ are normally distributed.
- Result 4.2: Each element in $\boldsymbol{X}$ follows a univariate normal distribution.
- Result 4.4: All subsets of the components of $\boldsymbol{X}$ have a multivariate normal distribution.
- Result 4.5: Zero covariance implies that the corresponding components are independently distributed.
- Result 4.6: The conditional distributions of the components are (multivariate) normal.
- Result 4.7: $\left(\boldsymbol{X}-\boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{X}-\boldsymbol{\mu}\right) \sim \chi^2_p$.
- Result 4.8: Linear combinations $\sum_{j=1}^n b_j \boldsymbol{X}_j$ and $\sum_{j=1}^n c_j \boldsymbol{X}_j$ are independent if $\displaystyle\sum_{j=1}^n c_j b_j = 0$.

Likelihood

- Suppose that $p \times 1$ vectors $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, ..., $\boldsymbol{X}_n$ represent a random sample from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The joint density function of all observations is

$$\prod_{j=1}^{n} \frac{1}{(2\pi)^{p/2}\sqrt{\det(\boldsymbol{\Sigma})}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}_j - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_j - \boldsymbol{\mu})\right\}.$$

- It is also called the likelihood, as a function of parameters, if we have observed values $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, ..., $\boldsymbol{x}_n$. We denote the likelihood by $L(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
- The estimator that maximize the likelihood is called the maximum likelihood estimator (MLE) and such estimation technique is called maximum likelihood estimation.

Likelihood of Normal Random Sample

We can write the likelihood as

$$
\begin{aligned}
&\prod_{j=1}^{n} \frac{1}{(2\pi)^{p/2}\sqrt{\det(\boldsymbol{\Sigma})}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}_j-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_j-\boldsymbol{\mu})\right\} \\
&= \exp\left\{-\frac{np}{2}\log(2\pi) - \frac{n}{2}\log\det(\boldsymbol{\Sigma}) - \frac{1}{2}\sum_{j=1}^{n}(\boldsymbol{x}_j-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_j-\boldsymbol{\mu})\right\}. \\
&= \exp\left\{-\frac{np}{2}\log(2\pi) - \frac{n}{2}\log\det(\boldsymbol{\Sigma}) - \frac{1}{2}\mathrm{tr}\left\{\boldsymbol{\Sigma}^{-1}\sum_{j=1}^{n}(\boldsymbol{x}_j-\boldsymbol{\mu})(\boldsymbol{x}_j-\boldsymbol{\mu})^T\right\}\right\}.
\end{aligned}
$$

The log-likelihood is

$$
\ell(\boldsymbol{\mu},\boldsymbol{\Sigma}) = \log L(\boldsymbol{\mu},\boldsymbol{\Sigma}).
$$

The MLE also maximizes the log-likelihood.

Derivation of MLE

Lemma: Optimization

1. Let $f(\boldsymbol{\mu}) = -\sum_{j=1}^{n} (\boldsymbol{x}_j - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_j - \boldsymbol{\mu})$. The maximum of $f(\boldsymbol{\mu})$ is attained when $\boldsymbol{\mu} = n^{-1} \sum_{j=1}^{n} \boldsymbol{x}_j$.
2. Let $g(\boldsymbol{\Sigma}) = -q \log\det(\boldsymbol{\Sigma}) - \operatorname{tr}\left\{\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\right\}$. The maximum of $g(\boldsymbol{\Sigma})$ is attained when $\boldsymbol{\Sigma} = q^{-1}\boldsymbol{A}$.

MLE

Result 4.11: Expression of MLE

Let $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, ..., $\boldsymbol{X}_n$ be a random sample from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then

$$\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{X}} \quad \text{and} \quad \hat{\boldsymbol{\Sigma}} = \frac{n-1}{n}\boldsymbol{S} = \boldsymbol{S}_n,$$

are the maximum likelihood estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively. Their observed values are the maximum likelihood estimates.

Note: the MLE of $\boldsymbol{\Sigma}$ is a biased estimator.

Invariance Property of MLE

Invariance Property of MLE

Let $\hat{\boldsymbol{\theta}}$ be the MLE of $\boldsymbol{\theta}$, and consider estimating the parameter $h(\boldsymbol{\theta})$ which is a function of $\boldsymbol{\theta}$. Then the MLE of $h(\boldsymbol{\theta})$ is $h\left(\hat{\boldsymbol{\theta}}\right)$.

Find MLE

Suppose that $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ are the MLEs of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively.

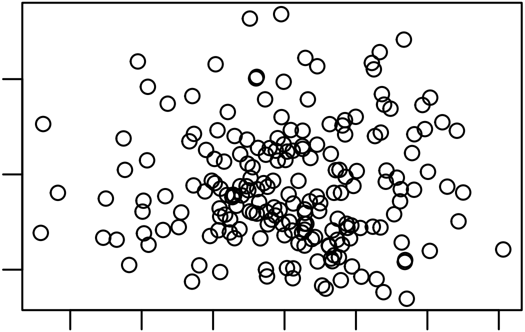
1. Find the MLE of $\boldsymbol{\mu}^T \boldsymbol{\Sigma} \boldsymbol{\mu}$.
2. Find the MLE of $\sqrt{\sigma_{ii}}$. Is the MLE an unbiased estimator?

Univariate Case

Let X_1, X_2, ..., X_n be a random sample from $N\left(\mu, \sigma^2\right)$. Then,

$$\bar{X} \sim N\left(\mu, \frac{1}{n}\sigma^2\right), \qquad \frac{(n-1)\,S^2}{\sigma^2} = \frac{\sum_{j=1}^{n}\left(X_j - \bar{X}\right)^2}{\sigma^2} \sim \chi^2_{n-1},$$

and $\bar{X}$ and S^2 are independent.



Chi-Square Distribution and Wishart Distribution

Let X_1, X_2, ..., X_m be mutually independent from $N(0,1)$. Then,

$$\sum_{j=1}^{m} X_j^2 \sim \chi_m^2.$$

Let $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, ..., $\boldsymbol{X}_n$ be mutually independent from $N_p(\mathbf{0},\boldsymbol{\Sigma})$. Then,

$$\sum_{j=1}^{m} \boldsymbol{X}_j \boldsymbol{X}_j^T \sim W_m(\cdot \mid \boldsymbol{\Sigma}),$$

a Wishart distribution.

It is also common to express a Wishart distribution as

$$\sum_{j=1}^{m} \boldsymbol{X}_j \boldsymbol{X}_j^T \sim W_p(\boldsymbol{\Sigma}, m),$$

where the expectation is $m\boldsymbol{\Sigma}$.

Properties of Wishart Distribution

1. If $\boldsymbol{A}_1 \sim W_p(\boldsymbol{\Sigma}, m_1)$ and $\boldsymbol{A}_2 \sim W_p(\boldsymbol{\Sigma}, m_2)$ independent of $\boldsymbol{A}_1$, then $\boldsymbol{A}_1 + \boldsymbol{A}_2 \sim W_p(\boldsymbol{\Sigma}, m_1 + m_2)$. That is, Wishart distribution is additive (so is chi square distribution).
2. If $\boldsymbol{A} \sim W_p(\boldsymbol{\Sigma}, m)$ and $\boldsymbol{C}$ is a $q \times p$ matrix of constants. Then, $\boldsymbol{C}\boldsymbol{A}\boldsymbol{C}^T \sim W_q(\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}^T, m)$.
3. If $\boldsymbol{A} \sim W_p(\boldsymbol{\Sigma}, m)$ and $\boldsymbol{c}$ is any fixed $p \times 1$ vector such that $\boldsymbol{c}^T\boldsymbol{\Sigma}\boldsymbol{c} \neq 0$. Then,

$$\frac{\boldsymbol{c}^T\boldsymbol{A}\boldsymbol{c}}{\boldsymbol{c}^T\boldsymbol{\Sigma}\boldsymbol{c}} \sim \chi^2_m.$$

This means that the marginal A_{ii} follows a scaled chi-square distribution.

Univariate to Multivariate

Let $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, ..., $\boldsymbol{X}_n$ be a random sample from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then

1. Result 4.8 indicates that

$$\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{X}} \quad \sim \quad N_p\left(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma}\right).$$

2. The distribution of

$$n\hat{\boldsymbol{\Sigma}} \quad = \quad n\boldsymbol{S}_n = \sum_{j=1}^{n}\left(\boldsymbol{X}_j - \bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_j - \bar{\boldsymbol{X}}\right)^T$$

   is $W_p(\boldsymbol{\Sigma}, n-1)$ (Cochran's Theorem).

3. $\bar{\boldsymbol{X}}$ and $\boldsymbol{S}_n$ are independent. In other words, $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ are independent.

Law of Large Numbers

Result 4.12: Weak Law of Large Numbers (LLN), Revised

Let $\boldsymbol{X}_1$, $\boldsymbol{X}$, ..., $\boldsymbol{X}_n$ be independent and identically distributed (i.i.d.) random vectors, and let $\boldsymbol{\mu} = \mathbb{E}(\boldsymbol{X}_1)$. If $\mathbb{E}\left(\sqrt{\boldsymbol{X}_1^T \boldsymbol{X}_1}\right) < \infty$, then

$$\bar{\boldsymbol{X}} = \frac{1}{n}\sum_{j=1}^{n} \boldsymbol{X}_j.$$

converges in probability to $\boldsymbol{\mu}$ as n increases. denoted by $\bar{\boldsymbol{X}} \xrightarrow{P} \boldsymbol{\mu}$. That is, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left(\sqrt{\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}\right)^T \left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}\right)} < \varepsilon\right) = 1.$$

Limit of MLE

1. By LLN,
$$\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{X}} \quad \sim \quad N_p\left(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma}\right),$$
converges in probability to $\boldsymbol{\mu}$.
2. For each sample covariance,
$$\begin{aligned} s_{ik} &= \frac{1}{n-1}\sum_{j=1}^{n}\left(X_{ji}-\bar{X}_i\right)\left(X_{jk}-\bar{X}_k\right) \\ &= \frac{1}{n-1}\sum_{j=1}^{n}\left(X_{ji}-\mu_i+\mu_i-\bar{X}_i\right)\left(X_{jk}-\mu_k+\mu_k-\bar{X}_k\right) \\ &= \frac{1}{n-1}\sum_{j=1}^{n}\left(X_{ji}-\mu_i\right)\left(X_{jk}-\mu_k\right)+\frac{n}{n-1}\left(\mu_i-\bar{X}_i\right)\left(\bar{X}_k-\mu_k\right) \end{aligned}$$
By LLN, s_{ik} converges in probability to σ_{jk}. And we say $\boldsymbol{S}$ (as well as $\boldsymbol{S}_n$) converges in probability to $\boldsymbol{\Sigma}$.

Central Limit Theorem

Result 4.13: Central Limit Theorem (CLT), Revised

Let $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, ..., $\boldsymbol{X}_n$ be independent and identically distributed random vectors from a population with mean $\boldsymbol{\mu}$ and finite covariance matrix $\boldsymbol{\Sigma}$. Then, $\boldsymbol{Z}_n = \sqrt{n}\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}\right)$ converges in distribution to $N_p\left(\boldsymbol{0}, \boldsymbol{\Sigma}\right)$ as n increases. That is, if $\boldsymbol{Z}_n$ converges in distribution to $\boldsymbol{Z}$, then

$$\lim_{n\to\infty} P\left(\boldsymbol{Z}_n \leq \boldsymbol{z}\right) = P\left(\boldsymbol{Z} \leq \boldsymbol{z}\right)$$

for all points $\boldsymbol{z}$ at which $P\left(\boldsymbol{Z} \leq \boldsymbol{z}\right)$ is continuous.

Distribution of MLE

1. Suppose that $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

$$\begin{aligned} \hat{\boldsymbol{\mu}} = \bar{\boldsymbol{X}} &\sim N_p\left(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma}\right), \\ n\hat{\boldsymbol{\Sigma}} = n\boldsymbol{S}_n &\sim W_p(\boldsymbol{\Sigma}, n-1), \\ n\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-1}\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}\right) &\sim \chi^2_p. \end{aligned}$$

   1. By LLN, $\boldsymbol{S}$ will be arbitrarily close to $\boldsymbol{\Sigma}$. And by CLT,

$$\begin{aligned} \sqrt{n}\boldsymbol{S}^{-1/2}\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}\right) &\text{ is approximately } N_p(\boldsymbol{0}, \boldsymbol{I}), \\ n\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}\right)^T \boldsymbol{S}^{-1}\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}\right) &\text{ is approximately } \chi^2_p, \end{aligned}$$

   for sufficiently large n.

2. By CLT, under mild conditions, $\sqrt{n}\boldsymbol{S}^{-1/2}\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}\right)$ is still approximately $N_p(\boldsymbol{0}, \boldsymbol{I})$ even if $\boldsymbol{X}$ is nonnormal.

Evaluate Univariate Normality

Commonly used methods include, but not limited to,

- Histogram, kernel density plot
- QQ plot
- Skewness 0 and kurtosis 3 (excess kurtosis 0)
- Shapiro-Wilk test
- Anderson-Darling test
- Kolmogorov-Smirnov test

Evaluate Multivariate Normality

Some methods to evaluate multivaraite normality are

- QQ plot (squared Mahalonobis distance against χ^2 distribution)
- Mardia's skewness and kurtosis
- Henze-Zirkler test
- Royston test: extension of Shapiro-Wilk test

Transformation to Near Normality

Some useful transformations are

Original Scale	Transformation
Counts	$\sqrt{y}$
Proportions	$\log\left(\frac{y}{1-y}\right)$
Correlation	$\log\left(\frac{1+y}{1-y}\right)$

Box-Cox transformation is

$$x^{(\lambda)} = \begin{cases} \frac{x^{\lambda}-1}{\lambda}, & \lambda \neq 0 \\ \log(x), & \lambda = 0 \end{cases}$$

for some cleverly chosen λ.