MULTIVARIATE NORMAL DISTRIBUTION

Reference: Johnson & Wichern(2007) *Applied Multivariate Statistical Analysis* Chapter 4.

1 Univariate Normal Distribution

We begin with the definition of univariate normal distribuion.

(a) Standard normal distribution $N(0,1)$. The probability density function $f(z)$ of a standard normal random variable Z is given by

$$f(z) = \frac{1}{\sqrt{2\pi}}\boldsymbol{e}^{-\frac{1}{2}z^2}\text{for} - \infty < z < \infty$$

Notation: $Z \sim N(0,1)$.

(b) Normal distribution with mean μ and variance $\sigma^2 N(\mu, \sigma^2)$. The probability density function(pdf) $f(x : \mu, \sigma^2)$ of the normal random variable with mean μ and variance σ^2 is

$$f(x : \mu, \sigma^2) = (2\pi)^{-\frac{1}{2}}(\sigma^2)^{-\frac{1}{2}}\boldsymbol{e}^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \text{ for } -\infty < x < \infty$$

Notation: $X \sim N(\mu, \sigma^2)$.

Note: $X = \sigma Z + \mu$.

2 Multivariate Normal Distribution

Let $\boldsymbol{Y}^T = (y_1 \; y_2 \; \dots \; y_p)$ be a random vector of p variate with mean vector $\boldsymbol{\mu}$ and covariance matrix σ. Now we can obtain the pdf of $\boldsymbol{Y}$ replacing x, μ and σ^2 in $f(x : \mu, \sigma^2)$ respectively by vectors $\boldsymbol{y}$, vector $\boldsymbol{\mu}$ and matrix Σ.

$$f_p(\boldsymbol{y} : \boldsymbol{\mu}, \Sigma) = (2\pi)^{-\frac{p}{2}}|\Sigma|^{-\frac{1}{2}}\boldsymbol{e}^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{y}-\boldsymbol{\mu})} \text{ for } \boldsymbol{y} \in \Re^p$$

where $\Re^p$ is the p-dimensional real space. Here $f_p(\boldsymbol{y} : \boldsymbol{\mu}, \Sigma)$ is called multivariate normal probability density function.

Notation: $\boldsymbol{Y} \sim N_p(\boldsymbol{\mu}, \Sigma)$.

Special Cases

(a) If $\boldsymbol{Y}$ has $N_p(\boldsymbol{\mu}, \Sigma)$, then the distribution of $\boldsymbol{Z} = (\boldsymbol{Y} - \boldsymbol{\mu})$ is $N_p(\mathbf{0}, \Sigma)$.

(b) If $\Sigma = \text{diag}(\ \sigma_{11} \quad \sigma_{22} \quad \ldots \quad \sigma_{pp}\ )$ then $Y_1, Y_2, \ldots, Y_p$ are independent normal random variables and

$$f_p(\boldsymbol{y} : \boldsymbol{\mu}, \Sigma) = f(y_1 : \mu_1, \sigma_{11}) f(y_2 : \mu_2, \sigma_{22}) \ldots f(y_p : \mu_p, \sigma_{pp})$$

(c) if $\boldsymbol{\mu} = \mathbf{0}$ and $\Sigma = \mathcal{I}$, then $Y_1, Y_2, \ldots, Y_p$, are independent identically distributed (standard) normal random variables.

(d) if $\boldsymbol{Z} = \mathcal{A}^{-1}(\boldsymbol{Y} - \boldsymbol{\mu})$ where $\mathcal{A}$ is a nonsingular matrix such that $\mathcal{A}\mathcal{A}^T = \Sigma$ then $\boldsymbol{Z} \sim N_p(\mathbf{0}, \mathcal{I})$. Note that $\boldsymbol{Y} = \mu + \mathcal{A}\boldsymbol{Z}$.

(e) Bivariate Normal Distribuion.

Consider a bivariate normal random vector $\boldsymbol{X}$. That is,

$$\boldsymbol{X} \sim N_2(\boldsymbol{\mu}, \Sigma)$$

where

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}.$$

Excercise: Compute the inverse of Σ and write down the probability density function of bivariate normal distribution.

3 Properties of the Multivariate Normal Distribution

(a) Suppose $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$, we have $\mathbf{E}(\boldsymbol{X}) = \boldsymbol{\mu}$, $\mathbf{Cov}(\boldsymbol{X}) = \Sigma$.

(b) Consider the transformations, $\boldsymbol{Z} = \mathcal{A}\boldsymbol{X} + \boldsymbol{b}$, where $\mathcal{A}_{p\times p}$ is a constant matrix,$\boldsymbol{b}_{q\times 1}$ is a constant vector, then

$$\boldsymbol{Z} \sim N_q(\mathcal{A}\boldsymbol{\mu} + \boldsymbol{b}, \mathcal{A}\Sigma\mathcal{A}^T).$$

(c) Sums of independent random vectors:

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_k$, be independent random vectors with distribuion $\boldsymbol{X}_j \sim N_p(\boldsymbol{\mu}_j, \Sigma)$ for $j = 1, 2, \ldots, k$ and $a_1, a_2, \ldots, a_k$ are real numbers. Define the random vector $\boldsymbol{Y}$ such that

$$\boldsymbol{Y} = a_1\boldsymbol{X}_1 + a_2\boldsymbol{X}_2 + \ldots + a_k\boldsymbol{X}_k = \sum_{j=1}^{k} a_j\boldsymbol{X}_j,$$

then

$$\boldsymbol{Y} \sim N_p\left(\sum_{j=1}^{k} a_j\boldsymbol{\mu}_j, \sum_{j=1}^{k} a_j^2\Sigma\right)$$

(d) Relation of the Multivariate Normal Distribution to χ^2 Distribuion:

For the univariate case, $X \sim N_1(\mu, \sigma^2)$, where $\sigma > 0$, we know that

$$Z^2 = \left\{\frac{X-\mu}{\sigma}\right\}^2 \sim \chi_1^2.$$

Similarly, suppose Σ is nonsingular and the random variable Y is given by

$$Y = (\boldsymbol{X} - \boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu})$$

then $Y \sim \chi_p^2$.

Proof: Since Σ is positive deinite and symmetric, there exist a non-singular matrix $\mathcal{B}$ such that $\Sigma = \mathcal{B}\mathcal{B}^T$.

Let $\boldsymbol{Z}$ be the random vector defined by $\boldsymbol{Z} = \mathcal{B}^{-1}(\boldsymbol{X} - \boldsymbol{\mu})$ then

$$\begin{aligned}
\boldsymbol{Y} &= (\boldsymbol{X} - \boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \\
&= (\boldsymbol{X} - \boldsymbol{\mu})^T(\mathcal{B}\mathcal{B}^T)^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \\
&= (\boldsymbol{X} - \boldsymbol{\mu})^T(\mathcal{B}^T)^{-1}(\mathcal{B})^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \\
&= (\mathcal{B}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}))^T(\mathcal{B}^{-1}(\boldsymbol{X} - \boldsymbol{\mu})) \\
&= \boldsymbol{Z}^T\boldsymbol{Z} = Z_1^2 + Z_2^2 + \ldots + Z_p^2
\end{aligned}$$

where $\boldsymbol{Z} = \begin{pmatrix} Z_1 & Z_2 \ldots & Z_p \end{pmatrix}$.

Since $\boldsymbol{Z} = \mathcal{B}^{-1}(\boldsymbol{X} - \boldsymbol{\mu})$, $\mathbf{E}(\boldsymbol{Z}) = \underset{\sim}{0}$ and

$$\begin{aligned}\mathbf{Cov}(\boldsymbol{Z}) &= \mathbf{Cov}(\mathcal{B}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})) = \mathcal{B}^{-1}\mathbf{Cov}(\boldsymbol{X}-\boldsymbol{\mu})(\mathcal{B}^{-1})^T \\ &= \mathcal{B}^{-1}\mathbf{Cov}(\boldsymbol{X})(\mathcal{B}^{-1})^T = \mathcal{B}^{-1}\Sigma(\mathcal{B}^{-1})^T \\ &= \mathcal{B}^{-1}(\mathcal{B}\mathcal{B}^T)(\mathcal{B}^{-1}) = (\mathcal{B}^{-1}\mathcal{B})(\mathcal{B}^T(\mathcal{B}^{-1})^T) = \mathcal{I}.\end{aligned}$$

Hence, $\boldsymbol{Z} \sim N_p(\mathbf{0}, \mathcal{I})$ and thus $Z_1, Z_2, \ldots, Z_p$, are $i.i.d$ $N(0,1)$. Therefore by univariate result $\boldsymbol{Y} = \sum_{j=1}^{p} Z_j^2 \sim \chi_p^2$.

4 Sampling Distributions

Consider a random sample $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ from a multivariate normal population with mean $\boldsymbol{\mu}$ and covariance Σ. Let $\overline{\boldsymbol{X}}$ be the sample mean vector and $\mathcal{S}_n$ be the sample covariance matrix computed using the above sample. Then

$$\overline{\boldsymbol{X}}_n = \frac{1}{n}\sum_{j=1}^{n}\boldsymbol{X}_j \text{ and } \mathcal{S}_n = \frac{1}{n-1}\sum_{j=1}^{n}(\boldsymbol{X}_j - \overline{\boldsymbol{X}}_n)(\boldsymbol{X}_j - \overline{\boldsymbol{X}}_n)^T.$$

(a) **Distribution of $\overline{X}_n$.**

Since $\mathbf{E}(\boldsymbol{X}_j) = \boldsymbol{\mu}$ and $\mathbf{Cov}(\boldsymbol{X}) = \Sigma$,

$$\mathbf{E}(\overline{\boldsymbol{X}}_n) = \boldsymbol{\mu} \text{ and } \mathbf{Cov}(\overline{\boldsymbol{X}}_n) = \frac{1}{n}\Sigma.$$

Further, since the population distribution is multivariate normal(that is, $\boldsymbol{X}_j$'s are multivariate normal random vectors,

$$\overline{\boldsymbol{X}}_n \sim N_p(\boldsymbol{\mu}, \frac{1}{n}\Sigma).$$

(b) **Distribuion of $\mathcal{S}_n$.**

$(n-1)\mathcal{S}_n$ has a Wishart distribution with $(n-1)$ degree of freedom(df). For more details of the Wishart distribution see Johnson & Wichern page 174.

Note: If $Z_1, Z_2, \ldots, Z_n$ are iid normal random vectors with $\mathbf{E}(\boldsymbol{Z}_i) = \mathbf{0}$ and $\mathbf{Cov}(\boldsymbol{Z}_i) = \Sigma$ for all i, then $\sum_{i=1}^{n} \boldsymbol{Z}_j \boldsymbol{Z}_j^T$ has a Wishart distribution with n df.

(c) **<u>Distribution of the Quatratic</u>**.

Let $Z^2 = n(\overline{\boldsymbol{X}}_n - \boldsymbol{\mu})^T \Sigma^{-1} (\overline{\boldsymbol{X}}_n - \boldsymbol{\mu})$. Then Z has a chi-square distribuion with p df.

(d) **<u>Independence of Sample Mean and Covariance Matrix</u>**.

$\overline{\boldsymbol{X}}_n$ and $\mathcal{S}_n$ are independent. Recall that in univariate case, the sample mean and the sample variance are independent. This is refer to as Basu's theorem.

(e) **<u>Hotelling T^2 Distribuion</u>**.

The distribuion of the statistics, T^2 where

$$T^2 = n(\overline{\boldsymbol{X}}_n - \boldsymbol{\mu})^T \mathcal{S}_n^{-1} (\overline{\boldsymbol{X}}_n - \boldsymbol{\mu}).$$

is called Hotelling's T^2 distribution.(For more details of the Hotelling's T^2 distribuion see Johnson & Wichern page 212.) Further we can prove that $\frac{(n-p)}{p(n-1)} T^2$ is distributed as $F_{p,n-p}$ (F distribution with p and $(n-p)df$).

(f) **<u>Mahalanobis Distance</u>**

Mahalanobis distance, D^2 is a standardised form of Euclidean distance. Mahalanobis distance of the obervation $\boldsymbol{X}_i$ in the sample given above is

$$D_i^2 = (\boldsymbol{X}_i - \overline{\boldsymbol{X}}_n)^T \mathcal{S}_n^{-1} (\boldsymbol{X}_i - \overline{\boldsymbol{X}}_n)$$

for $i = 1, 2, \ldots, n$.

If both n and $n - p$ are respectively greater than 30 and 25, then Mahalanobis distances, $D_i^2 (i = 1, 2, \ldots, n)$ are chi-square distributed with p degree of freedom.

5 Assessing the Normality of Multivariate Population

Let $\boldsymbol{X}$ be the random vector representing the multivariate population of interest and

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}$$

Theorem 1: if $\boldsymbol{X}$ is multivariate normal random vector then

(a) all the marginals are normally distributed, that is, X_j for all j, are normal random variables, and

(b) if a is a constant vector, then

$$\boldsymbol{a}^T\boldsymbol{X} \sim N(\boldsymbol{a}^T\boldsymbol{\mu}, \boldsymbol{a}^T\Sigma\boldsymbol{a})$$

where $\boldsymbol{\mu}$ is the mean and Σ is the covariance of $\boldsymbol{X}$.

Using the above themrem, we can prove the following theorem:

Theorem 2: A random vector $\boldsymbol{X}$ has a p-variate normal distribution if and only if (iff) $\boldsymbol{a}^T\boldsymbol{X}$ is normal for every constant vector $\boldsymbol{a}$.

Note:

(a) Theorem 2 is not use to prove that $\boldsymbol{X}$ is a normal random vector because it is difficult to prove that $\boldsymbol{a}^T\boldsymbol{X}$ is normal for all $\boldsymbol{a}$.

(b) Note that if $X_1, X_2, \ldots, X_p$ are normal random variables, it does not imply that $\boldsymbol{X}$ is multivariate normal. Prove this using the Example 4.8 page 202 in Johnson & Wichern.

6 Normal Probability Plot

Normal probability plot is also called a Q-Q(quantile quantile) plot.

(a) **Univariate Case**

Step 1: Arrange the observations $X_1, X_2, \ldots, X_n$ in increasing order. That is,

$$X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}.$$

Step 2: Obtain z_i from a Normal Table, such that $P(Z \leq z_i) = (i - 0.5)/n$ where $Z \sim N(0, 1)$.

Step 3: Plot the points $(z_i, X_{(i)}), i = 1, 2, \ldots, n$.

If the points approximately follow a straight line, then the observations are from a normal distribution.

(b) **Multivariate Case**

Step 1: Compute the Mahalanobis distances, D_i^2 of all the observations $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$,. Arrange the distances are in increasing order. That is,

$$D_{(1)}^2 \leq D_{(2)}^2 \leq \ldots \leq D_{(n)}^2.$$

Step 2: Obtain ξ_i from a Chi-Square Table, such that

$$\mathcal{P}(\chi_p^2 \leq \xi_i) = (i - 0.5)/n$$

for $i = 1, 2, \ldots, n$, where χ_p^2 is the chi-square distribution with p degrees of freedom.

Step 3: Plot the points $\left(\xi_i, D_{(i)}^2\right), i = 1, 2, \ldots, n$.

If the points approximately follow a straight line, then the observations are from a multivariate normal distribution.

7 Large Sample Distribution of $\overline{\boldsymbol{X}}_n$ and $\mathcal{S}_n$

Now we assume that the random sample, $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$, from a unknown multivariate population with mean $\boldsymbol{\mu}$ and covariance Σ.

(a) **Law of Large Numbers**.

1. $\overline{\boldsymbol{X}}_n$ converges in probability to $\boldsymbol{\mu}$: that is, $\mathcal{P}(|\overline{\boldsymbol{X}}_n - \boldsymbol{\mu}| < \epsilon) \to 1$ as $n \to \infty$ and
2. $\mathcal{S}_n$ converges in probability to Σ.

(b) **Central Limit Theorem**.

If n is large relative to p, then

$$\sqrt{n}(\overline{\boldsymbol{X}}_n - \boldsymbol{\mu}) \sim N_p(\mathbf{0}, \Sigma).$$

(c) **Important Results**.

If the sample covariance matrix is finite and non-singular and n is large relative to p, then the distribution of

- $\sqrt{n}(\overline{\boldsymbol{X}}_n - \boldsymbol{\mu})$ is approximately $N_p(\mathbf{0}, \Sigma)$ and
- $n(\overline{\boldsymbol{X}}_n - \boldsymbol{\mu})^T \mathcal{S}_n^{-1}(\overline{\boldsymbol{X}}_n - \boldsymbol{\mu})$ is approximately χ^2_p.