# Chapter 4.
# The Multivariate Normal Distribution.

## 4.1. Some properties about univariate normal distribution–a review.

Suppose $X \sim N(\mu, \sigma^2)$, a univariate normal distribution with mean $\mu$ and variance $\sigma^2$.

(i). Density:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad x \in (-\infty, \infty).$$

The density function is a bell-shaped curve, symmetric around the center $\mu$, with only one peak at $\mu$.

(ii). Linear transformation: Let $Y = aX + b$ Then, $Y \sim N(a\mu + b, a^2\sigma^2)$.

(iii). Linear combination of independent normal random variables is still normal. Suppose $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, ..., K$ are independent. Then, $\sum_{i=1}^K a_i X_i$ is still a normal random variable with mean $\sum_{i=1}^K a_i \mu_i$ and variance $\sum_{i=1}^K a_i^2 \sigma_i^2$.

(iv). Normal related distributions: $t$, $F$ and $\chi^2$: Suppose $\xi_i$ are iid $\sim N(0,1)$. Then, $t$, $F$ and $\chi^2$ distributions can be *defined* as follows

$$\sum_{i=1}^n \xi_i^2 \sim \chi_n^2, \qquad \frac{\xi_1}{\sqrt{[1/(n-1)] \sum_{j=2}^n \xi_j^2}} \sim t_{n-1}, \qquad \frac{(1/n)\sum_{i=1}^n \xi_i^2}{(1/m)\sum_{i=n+1}^{n+m} \xi_i^2} \sim F_{n,m}.$$

As a result, if $\xi_i$ are iid $\sim N(\mu, \sigma^2)$,

$$\sum_{i=1}^n (\xi_i - \bar{\xi}_n)^2/\sigma^2 \sim \chi_{n-1}^2, \qquad \frac{\bar{\xi}_n}{\sqrt{[1/(n-1)] \sum_{j=1}^n (\xi_j - \bar{\xi}_n)^2}} \sim t_{n-1},$$

$$\text{and} \qquad \frac{[1/(n-1)] \sum_{i=1}^n (\xi_i - \bar{\xi}_n)^2}{[1/(m-1)] \sum_{i=n+1}^{n+m} [\xi_i - (1/m)\sum_{k=n+1}^{n+m} \xi_k]^2} \sim F_{n-1,m-1}.$$

The proof uses a matrix transformation of the vector of $\xi_i$. (DIY).

(v). The central limit theorem: Suppose $\xi_i, i = 1, 2, ....$ are iid random variables with common mean $\mu$ and variance $\sigma^2$. Then,

$$\frac{\sum_{i=1}^n (\xi_i - \mu)}{\sqrt{n}\sigma} \to N(0,1), \qquad \text{in distribution,}$$

$$i.e., \qquad \frac{\bar{\xi}_n - \mu}{\sigma/\sqrt{n}} \to N(0,1), \qquad \text{in distribution.}$$

where $\bar{\xi}_n = (1/n)\sum_{i=1}^n \xi_i$. Moreover,

$$\frac{\sqrt{n}(\bar{\xi}_n - \mu)}{\sqrt{1/(n-1)\sum_{j=1}^n (\xi_j - \bar{\xi}_n)^2}} \to N(0,1), \qquad \text{in distribution.}$$

## 4.2. Multivariate normal distribution.

Suppose

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}_{p\times 1} \sim MN\Big(\mu_{p\times 1}, \Sigma_{p\times p}\Big), \qquad \text{where} \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}_{p\times 1}, \ \Sigma = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & \vdots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{pmatrix}_{p\times p}$$

and $\sigma_{ij}$ =cov$(X_i, X_j)$. Then,

(i). Marginal normality: $X_i \sim N(\mu_i, \sigma_{ii})$. Multivariate normality implies marginal normality.

(ii). Transformation invariant: $AX \sim MN(A\mu, A\Sigma A')$ where $A$ is any (nonrandom) $r \times p$ matrix. Linear transformation of multivariate normal random variable is still multivariate normal.

(iii). Zero correlation is equivalent to independence: $X_1, ..., X_p$ are independent if and only if $\sigma_{ij} = 0$ for $1 \le i \ne j \le p$. Or, in other words, if and only if $\Sigma$ is diagonal.

(iv). Standardization as a special linear transformation:

$$\Sigma^{-1/2}(X - \mu) \sim MN(0, I_p).$$

Here, 0 is a $p$-vector with all $p$ entries being 0, $I_p$ shall always denote the identity $p \times p$ matrix, $\Sigma^{-1/2}$ is any matrix such that $\Sigma^{-1/2}\Sigma(\Sigma^{-1/2})' = I_p$.

Remark. $\Sigma^{-1/2}$ is not uniquely defined. Suppose $\Sigma = T\Lambda T'$ where $T$ is orthonormal, i.e., $TT' = I_p$, and $\Lambda$ diagonal. Then, $\Sigma^{-1/2}$ can be expressed as $\tilde{T}\Lambda^{-1/2}T'$ where $\tilde{T}$ can be any orthonormal matrix. Unless otherwise stated, $\Sigma^{-1/2}$ and $\Sigma^{1/2}$ shall be the symmetric ones.

(v). Conditional distribution:

**Proposition 4.1.** *Suppose*

$$X_{(p+q)\times 1} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim MN\Big(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{11} & \Sigma_{12} \end{pmatrix}\Big),$$

*where $X_1$ and $\mu_1$ are $p$-vectors, $X_2$ and $\mu_2$ are $q$-vectors, and $\Sigma_{11}, \Sigma_{12}, \Sigma_{21}$ and $\Sigma_{22}$ are $p \times p$, $p \times q$, $q \times p$ and $q \times q$ matrices, respectively. Then, the conditional distribution of $X_1$ given $X_2 = x_2$ is*

$$MN(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2),\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}).$$

Proof. Observe that

$$X_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2) \qquad \text{and} \quad X_2 - \mu_2$$

are 0-correlated. Therefore, the two random vectors are independent of each other. Hence the conditional distribution of the former given $X_2 = x_2$, i.e, $X_2 - \mu_2 = x_2 - \mu_2$, follows the same distribution as the unconditional one, which is $MN(0, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}')$. This implies $X_1$ given $X_2 = x_2$ follows $MN(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$. $\square$

Remark. The above proposition has important connection with the geometric structure of mean 0 random variables with applications in linear regression. For simplicity, consider linear regression of $Y$, a response of 1 dimension, on $X$, a covariate random variable of 1 dimension:

$$Y = \alpha + \beta X + \epsilon,$$

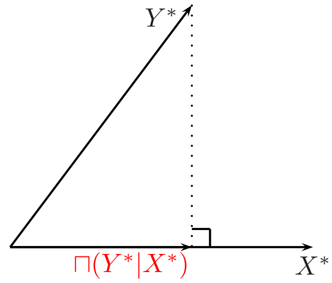
with $\epsilon$ being mean 0 and uncorrelated with $X$, which implies $Y - \alpha - \beta X$ is mean 0 and uncorrelated with $X$. Then $\alpha = \mu_Y + \beta\mu_X$ in order for $Y - \alpha - \beta X$ to be mean 0, and, as a result, $\beta$ is identified as the constant such that

$$Y - \mu_Y - \beta(X - \mu_X) \perp (X - \mu_X).$$

Here $\perp$ means 0-correlated. Now picture $Y^* = Y - \mu_Y$ and $X^* = X - \mu_X$ as two vectors in a space of mean 0 finite variance random variables equipped with inner product, the covariance of the two random variables. The projection of $Y^*$ on $X^*$, denoted as $\sqcap(Y^*|X^*)$, is then

$$\sqcap(Y^*|X^*) = \frac{< Y^*, X^* >}{\|X^*\|^2} X^* = \frac{cov(Y, X)}{var(X)}(X - \mu_X).$$

This is verified by checking that $< Y^* - \sqcap(Y^*|X^*),\ X^* >= 0$.




As a result, $\beta = cov(Y, X)/var(X)$ in the linear regression of $Y$ on $X$, which gives rise to the least squares estimator

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

(vi). Density function:

$$f(x) = \frac{1}{\sqrt{(2\pi)^p|\Sigma|}} \exp\{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\}, \qquad x \in R^p.$$

The constant density contour is defined as

$$\{x : f(x) = a\} = \{x : (x-\mu)'\Sigma^{-1}(x-\mu) = c\},$$

which is an ellipse, where $c = -2[\log a + (1/2)\log((2\pi)^p|\Sigma|)]$. Note that

$$(X-\mu)'\Sigma^{-1}(X-\mu) \sim \chi_p^2,$$

since $\Sigma^{-1/2}(X-\mu) \sim MN(0, I_p)$.