Lecture 1. Random vectors and multivariate normal distribution

1.1 Moments of random vector

A random vector \( \mathbf{X} \) of size \( p \) is a column vector consisting of \( p \) random variables \( X_1, \ldots, X_p \) and is \( \mathbf{X} = (X_1, \ldots, X_p)' \). The mean or expectation of \( \mathbf{X} \) is defined by the vector of expectations,

\[
\mu \equiv E(\mathbf{X}) = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_p) \end{pmatrix},
\]

which exists if \( E|X_i| < \infty \) for all \( i = 1, \ldots, p \).

Lemma 1. Let \( \mathbf{X} \) be a random vector of size \( p \) and \( \mathbf{Y} \) be a random vector of size \( q \). For any non-random matrices \( \mathbf{A}_{(m \times p)}, \mathbf{B}_{(m \times q)}, \mathbf{C}_{(1 \times n)}, \) and \( \mathbf{D}_{(m \times n)} \),

\[
E(\mathbf{AX} + \mathbf{BY}) = \mathbf{A}E(\mathbf{X}) + \mathbf{B}E(\mathbf{Y}),
\]

\[
E(\mathbf{AXC} + \mathbf{D}) = \mathbf{A}E(\mathbf{X})\mathbf{C} + \mathbf{D}.
\]

For a random vector \( \mathbf{X} \) of size \( p \) satisfying \( E(X_i^2) < \infty \) for all \( i = 1, \ldots, p \), the variance-covariance matrix (or just covariance matrix) of \( \mathbf{X} \) is

\[
\Sigma \equiv \text{Cov}(\mathbf{X}) = E[(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})'].
\]

The covariance matrix of \( \mathbf{X} \) is a \( p \times p \) square, symmetric matrix. In particular, \( \Sigma_{ij} = \text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i) = \Sigma_{ji} \).

Some properties:

1. \( \text{Cov}(\mathbf{X}) = E(\mathbf{XX}') - E(\mathbf{X})E(\mathbf{X})' \).
2. If \( \mathbf{c} = \mathbf{c}_{(p \times 1)} \) is a constant, \( \text{Cov}(\mathbf{X} + \mathbf{c}) = \text{Cov}(\mathbf{X}) \).
3. If \( \mathbf{A}_{(m \times p)} \) is a constant, \( \text{Cov}(\mathbf{AX}) = \mathbf{A}\text{Cov}(\mathbf{X})\mathbf{A}' \).

Lemma 2. The \( p \times p \) matrix \( \Sigma \) is a covariance matrix if and only if it is non-negative definite.

1.2 Multivariate normal distribution - nonsingular case

Recall that the univariate normal distribution with mean \( \mu \) and variance \( \sigma^2 \) has density

\[
f(x) = (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{1}{2}(x - \mu)\sigma^{-2}(x - \mu)\right].
\]

Similarly, the multivariate normal distribution for the special case of nonsingular covariance matrix \( \Sigma \) is defined as follows.
Definition 1. Let \( \mu \in \mathbb{R}^p \) and \( \Sigma_{(p \times p)} > 0 \). A random vector \( X \in \mathbb{R}^p \) has \( p \)-variate normal distribution with mean \( \mu \) and covariance matrix \( \Sigma \) if it has probability density function

\[
f(x) = |2\pi \Sigma|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2}(x - \mu)'\Sigma^{-1}(x - \mu) \right],
\]

for \( x \in \mathbb{R}^p \). We use the notation \( X \sim N_p(\mu, \Sigma) \).

Theorem 3. If \( X \sim N_p(\mu, \Sigma) \) for \( \Sigma > 0 \), then

1. \( Y = \Sigma^{-\frac{1}{2}}(X - \mu) \sim N_p(0, I_p) \),
2. \( X = \Sigma^{\frac{1}{2}}Y + \mu \) where \( Y \sim N_p(0, I_p) \),
3. \( E(X) = \mu \) and \( \text{Cov}(X) = \Sigma \),
4. for any fixed \( v \in \mathbb{R}^p \), \( v'X \) is univariate normal.
5. \( U = (X - \mu)'\Sigma^{-1}(X - \mu) \sim \chi^2(p) \).

Example 1 (Bivariate normal).

1.2.1 Geometry of multivariate normal

The multivariate normal distribution has location parameter \( \mu \) and the shape parameter \( \Sigma > 0 \). In particular, let's look into the contour of equal density

\[
E_c = \{ x \in \mathbb{R}^p : f(x) = c_0 \}
= \{ x \in \mathbb{R}^p : (x - \mu)'\Sigma^{-1}(x - \mu) = c^2 \}.
\]

Moreover, consider the spectral decomposition of \( \Sigma = U\Lambda U' \) where \( U = [u_1, \ldots, u_p] \) and \( \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_p) \) with \( \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p > 0 \). The \( E_c \), for any \( c > 0 \), is an ellipsoid centered around \( \mu \) with principal axes \( u_i \) of length proportional to \( \sqrt{\lambda_i} \). If \( \Sigma = I_p \), the ellipsoid is the surface of a sphere of radius \( c \) centered at \( \mu \).

As an example, consider a bivariate normal distribution \( N_2(0, \Sigma) \) with

\[
\Sigma = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} \cos(\pi/4) & -\sin(\pi/4) \\ \sin(\pi/4) & \cos(\pi/4) \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos(\pi/4) & -\sin(\pi/4) \\ \sin(\pi/4) & \cos(\pi/4) \end{bmatrix}'.
\]

The location of the distribution is the origin \( (\mu = 0) \), and the shape \( (\Sigma) \) of the distribution is determined by the ellipse given by the two principal axes (one at 45 degree line, the other at -45 degree line). Figure 1 shows the density function and the corresponding \( E_c \) for \( c = 0.5, 1, 1.5, 2, \ldots \).
1.3 General multivariate normal distribution

The characteristic function of a random vector $X$ is defined as

$$\varphi_X(t) = E(e^{it'X}), \quad \text{for} \ t \in \mathbb{R}^p.$$ 

Note that the characteristic function is $\mathbb{C}$-valued, and always exists. We collect some important facts.

1. $\varphi_X(t) = \varphi_Y(t)$ if and only if $X \overset{d}{=} Y$.
2. If $X$ and $Y$ are independent, then $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$.
3. $X_n \Rightarrow X$ if and only if $\varphi_{X_n}(t) \to \varphi_X(t)$ for all $t$.

An important corollary follows from the uniqueness of the characteristic function.

**Corollary 4** (Cramer–Wold device). *If $X$ is a $p \times 1$ random vector then its distribution is uniquely determined by the distributions of linear functions of $t'X$, for every $t \in \mathbb{R}^p$.*

Corollary 4 paves the way to the definition of (general) multivariate normal distribution.

**Definition 2.** A random vector $X \in \mathbb{R}^p$ has a multivariate normal distribution if $t'X$ is an univariate normal for all $t \in \mathbb{R}^p$.

The definition says that $X$ is MVN if every projection of $X$ onto a 1-dimensional subspace is normal, with a convention that a degenerate distribution $\delta_c$ has a normal distribution with variance 0, i.e., $c \sim N(c, 0)$. The definition does not require that $\text{Cov}(X)$ is nonsingular.
Theorem 5. The characteristic function of a multivariate normal distribution with mean $\mu$ and covariance matrix $\Sigma \geq 0$ is, for $t \in \mathbb{R}^p$, 

$$\varphi(t) = \exp[i t' \mu - \frac{1}{2} t' \Sigma t].$$

If $\Sigma > 0$, then the pdf exists and is the same as (1).

In the following, the notation $X \sim N(\mu, \Sigma)$ is valid for a non-negative definite $\Sigma$. However, whenever $\Sigma^{-1}$ appears in the statement, $\Sigma$ is assumed to be positive definite.

Proposition 6. If $X \sim N_p(\mu, \Sigma)$ and $Y = AX + b$ for $A_{(q \times p)}$ and $b_{(q \times 1)}$, then $Y \sim N_q(A\mu + b, A\Sigma A')$.

Next two results are concerning independence and conditional distributions of normal random vectors. Let $X_1$ and $X_2$ be the partition of $X$ whose dimensions are $r$ and $s$, $r + s = p$, and suppose $\mu$ and $\Sigma$ are partitioned accordingly. That is,

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N_p \left( \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right).$$

Proposition 7. The normal random vectors $X_1$ and $X_2$ are independent if and only if $\text{Cov}(X_1, X_2) = \Sigma_{12} = 0$.

Proposition 8. The conditional distribution of $X_1$ given $X_2 = x_2$ is

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

Proof. Consider new random vectors $X_1^* = X_1 - \Sigma_{12} \Sigma_{22}^{-1} X_2$ and $X_2^* = X_2$,

$$X^* = \begin{bmatrix} X_1^* \\ X_2^* \end{bmatrix} = AX, \quad A = \begin{bmatrix} \mathbb{I}_r & -\Sigma_{12} \Sigma_{22}^{-1} \\ 0_{(s \times r)} & \mathbb{I}_s \end{bmatrix}.$$ 

By Proposition 6, $X^*$ is multivariate normal. An inspection of the covariance matrix of $X^*$ leads that $X_1^*$ and $X_2^*$ are independent. The result follows by writing

$$X_1 = X_1^* + \Sigma_{12} \Sigma_{22}^{-1} X_2,$$

and that the distribution (law) of $X_1$ given $X_2 = x_2$ is $\mathcal{L}(X_1 \mid X_2 = x_2) = \mathcal{L}(X_1^* + \Sigma_{12} \Sigma_{22}^{-1} x_2 \mid X_2 = x_2)$, which is a MVN of dimension $r$. \hfill $\square$
1.4 Multivariate Central Limit Theorem

If $X_1, X_2, \ldots \in \mathbb{R}^p$ are i.i.d. with $E(X_i) = \mu$ and $\text{Cov}(X) = \Sigma$, then

$$n^{-\frac{1}{2}} \sum_{j=1}^{n} (X_j - \mu) \Rightarrow N_p(0, \Sigma) \quad \text{as} \quad n \to \infty,$$

or equivalently,

$$n^{\frac{1}{2}}(\bar{X}_n - \mu) \Rightarrow N_p(0, \Sigma) \quad \text{as} \quad n \to \infty,$$

where $\bar{X}_n = \frac{1}{n} \sum_{j=1}^{n} X_j$.

The delta-method can be used for asymptotic normality of $h(\bar{X}_n)$ for some function $h: \mathbb{R}^p \to \mathbb{R}$. In particular, denote $\nabla h(x)$ for the gradient of $h$ at $x$. Using the first two terms of Taylor series,

$$h(\bar{X}_n) = h(\mu) + (\nabla h(\mu))'(\bar{X}_n - \mu) + O_p(\|\bar{X}_n - \mu\|^2),$$

Then Slutsky’s theorem gives the result,

$$\sqrt{n}(h(\bar{X}_n) - h(\mu)) = (\nabla h(\mu))' \sqrt{n}(\bar{X}_n - \mu) + O_p(\sqrt{n}(\bar{X}_n - \mu)'(\bar{X}_n - \mu))$$

$$\Rightarrow (\nabla h(\mu))' N_p(0, \Sigma) \quad \text{as} \quad n \to \infty,$$

$$= N_p(0, (\nabla h(\mu))' \Sigma(\nabla h(\mu))).$$

1.5 Quadratic forms in normal random vectors

Let $X \sim N_p(\mu, \Sigma)$. A quadratic form in $X$ is a random variable of the form

$$Y = X'AX = \sum_{i=1}^{p} \sum_{j=1}^{p} X_i a_{ij} X_j,$$

where $A$ is a $p \times p$ symmetric matrix and $X_i$ is the $i$th element of $X$. We are interested in the distribution of quadratic forms and the conditions under which two quadratic forms are independent.

Example 2. A special case: If $X \sim N_p(0, I_p)$ and $A = I_p$,

$$Y = X'AX = X'X = \sum_{i=1}^{p} X_i^2 \sim \chi^2(p).$$

Fact 1. Recall the following:

1. A $p \times p$ matrix $A$ is idempotent if $A^2 = A$.
2. If $A$ is symmetric, then $A = \Gamma' \Lambda \Gamma$, where $\Lambda = \text{diag}(\lambda_i)$ and $\Gamma$ is orthogonal.
3. If $A$ is symmetric idempotent,
(a) its eigenvalues are either 0 or 1,
(b) \( \text{rank}(A) = \#\{\text{non zero eigenvalues}\} = \text{trace}(A) \).

**Theorem 9.** Let \( X \sim N_p(0, \sigma^2 \mathbb{I}) \) and \( A \) be a \( p \times p \) symmetric matrix. Then
\[
Y = \frac{X'AX}{\sigma^2} \sim \chi^2(m)
\]
if and only if \( A \) is idempotent of rank \( m < p \).

**Corollary 10.** Let \( X \sim N_p(0, \Sigma) \) and \( A \) be a \( p \times p \) symmetric matrix. Then
\[
Y = X'AX \sim \chi^2(m)
\]
if and only if either i) \( A\Sigma \) is idempotent of rank \( m \) or ii) \( \Sigma A \) is idempotent of rank \( m \).

**Example 3.** If \( X \sim N_p(\mu, \Sigma) \) then \( (X - \mu)'\Sigma^{-1}(X - \mu) \sim \chi^2(p) \).

**Theorem 11.** Let \( X \sim N_p(0, \mathbb{I}) \) and \( A \) be a \( p \times p \) symmetric matrix, and \( B \) be a \( k \times p \) matrix. If \( BA = 0 \), then \( BX \) and \( X'AX \) are independent.

**Example 4.** Let \( X_i \sim N(\mu, \sigma^2) \) i.i.d. The sample mean \( \bar{X}_n \) and the sample variance \( S_n^2 = (n - 1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \) are independent. Moreover, \( (n - 1)\frac{S_n^2}{\sigma^2} \sim \chi^2(n - 1) \).

**Theorem 12.** Let \( X \sim N_p(0, \mathbb{I}) \). Suppose \( A \) and \( B \) are \( p \times p \) symmetric matrices. If \( BA = 0 \), then \( X'AX \) and \( X'BX \) are independent.

**Corollary 13.** Let \( X \sim N_p(0, \Sigma) \) and \( A \) be a \( p \times p \) symmetric matrix.

1. For \( B_{(k \times p)} \), \( BX \) and \( X'AX \) are independent if \( B\Sigma A = 0 \);
2. For symmetric \( B \), \( X'AX \) and \( X'BX \) are independent if \( B\Sigma A = 0 \).

**Example 5.** The residual sum of squares in the standard linear regression has a scaled chi-squared distribution and is independent with the coefficient estimates.

Next lecture is on the distribution of the sample covariance matrix.