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,

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

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

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

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

Some properties:

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

Lemma 2. *The $p \times p$ matrix Σ 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 μ and variance σ^2 has density

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

Similarly, the multivariate normal distribution for the special case of nonsingular covariance matrix Σ is defined as follows.

Definition 1. Let $\mu \in \mathbb{R}^p$ and $\Sigma_{(p\times p)} > 0$. A random vector $\mathbf{X} \in \mathbb{R}^p$ has p-variate normal distribution with mean μ and covariance matrix Σ if it has probability density function

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

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

Theorem 3. *If $\mathbf{X} \sim N_p(\mu, \Sigma)$ for $\Sigma > 0$, then*

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

Example 1 (Bivariate normal).

1.2.1 Geometry of multivariate normal

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

$$\begin{aligned} E_c &= \{\mathbf{x} \in \mathbb{R}^p : f(\mathbf{x}) = c_0\} \\ &= \{\mathbf{x} \in \mathbb{R}^p : (\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu) = c^2\}. \end{aligned}$$

Moreover, consider the spectral decomposition of $\Sigma = \mathbf{U}\Lambda\mathbf{U}'$ where $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_p]$ 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 μ with principal axes $\mathbf{u}_i$ of length proportional to $\sqrt{\lambda_i}$. If $\Sigma = \mathbb{I}_p$, the ellipsoid is the surface of a sphere of radius c centered at μ.

As an example, consider a bivariate normal distribution $N_2(\mathbf{0}, \Sigma)$ 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 ($\mu = \mathbf{0}$), and the shape (Σ) 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 $c = 0.5, 1, 1.5, 2, \ldots$.

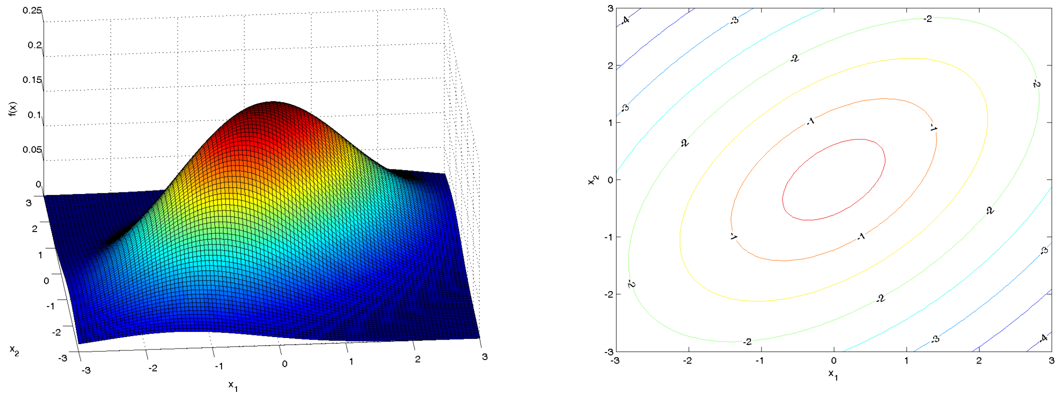

Figure 1: Bivariate normal density and its contours. Notice that an ellipses in the plane can represent a bivariate normal distribution. In higher dimensions $d > 2$, ellipsoids play the similar role.

1.3 General multivariate normal distribution

The characteristic function of a random vector $\mathbf{X}$ is defined as

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

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

1. $\varphi_{\mathbf{X}}(t) = \varphi_{\mathbf{Y}}(t)$ if and only if $\mathbf{X} \stackrel{\mathcal{L}}{=} \mathbf{Y}$.
2. If $\mathbf{X}$ and $\mathbf{Y}$ are independent, then $\varphi_{\mathbf{X}+\mathbf{Y}} = \varphi_{\mathbf{X}}(t)\varphi_{\mathbf{Y}}(t)$.
3. $\mathbf{X}_n \Rightarrow \mathbf{X}$ if and only if $\varphi_{\mathbf{X}_n}(t) \to \varphi_{\mathbf{X}}(t)$ for all t.

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

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

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

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

The definition says that $\mathbf{X}$ is MVN if every projection of $\mathbf{X}$ onto a 1-dimensional subspace is normal, with a convention that a degenerate distribution δ_c has a normal distribution with variance 0, i.e., $c \sim N(c, 0)$. The definition does not require that $\text{Cov}(\mathbf{X})$ is nonsingular.

Theorem 5. *The characteristic function of a multivariate normal distribution with mean μ and covariance matrix $\Sigma \geq 0$ is, for $t \in \mathbb{R}^p$,*

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

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

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

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

Next two results are concerning independence and conditional distributions of normal random vectors. Let $\mathbf{X}_1$ and $\mathbf{X}_2$ be the partition of $\mathbf{X}$ whose dimensions are r and s, $r + s = p$, and suppose μ and Σ are partitioned accordingly. That is,

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \sim N_p\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right).$$

Proposition 7. *The normal random vectors $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent if and only if $Cov(\mathbf{X}_1, \mathbf{X}_2) = \Sigma_{12} = \mathbf{0}$.*

Proposition 8. *The conditional distribution of $\mathbf{X}_1$ given $\mathbf{X}_2 = \mathbf{x}_2$ is*

$$N_r(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

Proof. Consider new random vectors $\mathbf{X}_1^* = \mathbf{X}_1 - \Sigma_{12}\Sigma_{22}^{-1}\mathbf{X}_2$ and $\mathbf{X}_2^* = \mathbf{X}_2$,

$$\mathbf{X}^* = \begin{bmatrix} \mathbf{X}_1^* \\ \mathbf{X}_2^* \end{bmatrix} = \mathbf{A}\mathbf{X}, \quad \mathbf{A} = \begin{bmatrix} \mathbb{I}_r & -\Sigma_{12}\Sigma_{22}^{-1} \\ \mathbf{0}_{(s\times r)} & \mathbb{I}_s \end{bmatrix}.$$

By Proposition 6, $\mathbf{X}^*$ is multivariate normal. An inspection of the covariance matrix of $\mathbf{X}^*$ leads that $\mathbf{X}_1^*$ and $\mathbf{X}_2^*$ are independent. The result follows by writing

$$\mathbf{X}_1 = \mathbf{X}_1^* + \Sigma_{12}\Sigma_{22}^{-1}\mathbf{X}_2,$$

and that the distribution (law) of $\mathbf{X}_1$ given $\mathbf{X}_2 = \mathbf{x}_2$ is $\mathcal{L}(\mathbf{X}_1 \mid \mathbf{X}_2 = \mathbf{x}_2) = \mathcal{L}(\mathbf{X}_1^* + \Sigma_{12}\Sigma_{22}^{-1}\mathbf{X}_2 \mid \mathbf{X}_2 = \mathbf{x}_2) = \mathcal{L}(\mathbf{X}_1^* + \Sigma_{12}\Sigma_{22}^{-1}\mathbf{x}_2 \mid \mathbf{X}_2 = \mathbf{x}_2)$, which is a MVN of dimension r. □

1.4 Multivariate Central Limit Theorem

If $\mathbf{X}_1, \mathbf{X}_2, \ldots \in \mathbb{R}^p$ are i.i.d. with $E(\mathbf{X}_i) = \mu$ and $\mathrm{Cov}(\mathbf{X}) = \Sigma$, then

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

or equivalently,

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

where $\bar{\mathbf{X}}_n = \frac{1}{2} \sum_{j=1}^{n} \mathbf{X}_j$.

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

$$h(\bar{\mathbf{X}}_n) = h(\mu) + (\nabla h(\mu))'(\bar{\mathbf{X}}_n - \mu) + O_p(\|\bar{\mathbf{X}}_n - \mu\|_2^2),$$

Then Slutsky's theorem gives the result,

$$\begin{aligned} \sqrt{n}(h(\bar{\mathbf{X}}_n) - h(\mu)) &= (\nabla h(\mu))' \sqrt{n}(\bar{\mathbf{X}}_n - \mu) + O_p(\sqrt{n}(\bar{\mathbf{X}}_n - \mu)'(\bar{\mathbf{X}}_n - \mu)) \\ &\Rightarrow (\nabla h(\mu))' N_p(\mathbf{0}, \Sigma) \quad \text{as} \;\; n \to \infty, \\ &= N_p(\mathbf{0}, (\nabla h(\mu))' \Sigma (\nabla h(\mu))) \end{aligned}$$

1.5 Quadratic forms in normal random vectors

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

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

where $\mathbf{A}$ is a $p \times p$ symmetric matrix. 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 $\mathbf{X} \sim N_p(0, \mathbb{I}_p)$ and $\mathbf{A} = \mathbb{I}_p$,

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

Fact 1. Recall the following:

1. A $p \times p$ matrix $\mathbf{A}$ is idempotent if $\mathbf{A}^2 = \mathbf{A}$.
2. If $\mathbf{A}$ is symmetric, then $\mathbf{A} = \Gamma' \Lambda \Gamma$, where $\Lambda = \mathrm{diag}(\lambda_i)$ and Γ is orthogonal.
3. If $\mathbf{A}$ is symmetric idempotent,
    (a) its eigenvalues are either 0 or 1,

(b) $\operatorname{rank}(\mathbf{A}) = \#\{\text{non zero eigenvalues}\} = \operatorname{trace}(\mathbf{A})$.

Theorem 9. *Let $\mathbf{X} \sim N_p(0, \sigma^2 \mathbb{I})$ and $\mathbf{A}$ be a $p \times p$ symmetric matrix. Then*

$$Y = \frac{\mathbf{X}'\mathbf{A}\mathbf{X}}{\sigma^2} \sim \chi^2(m)$$

if and only if $\mathbf{A}$ is idempotent of rank $m < p$.

Corollary 10. *Let $\mathbf{X} \sim N_p(0, \Sigma)$ and $\mathbf{A}$ be a $p \times p$ symmetric matrix. Then*

$$Y = \mathbf{X}'\mathbf{A}\mathbf{X} \sim \chi^2(m)$$

if and only if either i) $\mathbf{A}\Sigma$ is idempotent of rank m or ii) $\Sigma\mathbf{A}$ is idempotent of rank m.

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

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

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

1. *For $\mathbf{B}_{(k \times p)}$, $\mathbf{B}\mathbf{X}$ and $\mathbf{X}'\mathbf{A}\mathbf{X}$ are independent if $\mathbf{B}\Sigma\mathbf{A} = 0$;*
2. *For symmetric $\mathbf{B}$, $\mathbf{X}'\mathbf{A}\mathbf{X}$ and $\mathbf{X}'\mathbf{B}\mathbf{X}$ are independent if $\mathbf{B}\Sigma\mathbf{A} = 0$.*

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.