15 of August 2015

Al-Ahmadgaid B. Asaad
*Lecture 1* | **Stat233**

# 1 Basics

1. Idempotent matrix must necessarily be square matrix.
2.

# 2 Subspaces and Projections

1. Inner Product

   **Definition 1.** Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. The inner product of $\mathbf{u}$ and $\mathbf{v}$ denoted by $\langle \mathbf{u}, \mathbf{v} \rangle$ is given by

   $$\mathbf{u}'\mathbf{v} = \sum_{i=1}^{n} u_i v_i.$$

   (a) $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$

   $$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^{n} u_i v_i = \sum_{i=1}^{n} v_i u_i = \mathbf{v}'\mathbf{u} = \langle \mathbf{v}, \mathbf{u} \rangle$$

   (b) $\langle a\mathbf{u}, \mathbf{v} \rangle = a\langle \mathbf{u}, \mathbf{v} \rangle$

   $$\langle a\mathbf{u}, \mathbf{v} \rangle = a\mathbf{u}'\mathbf{v} = \sum_{i=1}^{n} a u_i v_i = a \sum_{i=1}^{n} u_i v_i = a\langle \mathbf{u}, \mathbf{v} \rangle$$

   (c) $\langle \mathbf{u}_1 + \mathbf{u}_2, \mathbf{v} \rangle = \langle \mathbf{u}_1, \mathbf{v} \rangle + \langle \mathbf{u}_2, \mathbf{v} \rangle$

   $$\langle \mathbf{u}_1 + \mathbf{u}_2, \mathbf{v} \rangle = (\mathbf{u}_1 + \mathbf{u}_2)'\mathbf{v} = \mathbf{u}_1'\mathbf{v} + \mathbf{u}_2'\mathbf{v} = \langle \mathbf{u}_1, \mathbf{v} \rangle + \langle \mathbf{u}_2, \mathbf{v} \rangle$$

2. Linear Combinations

   **Definition 2.** Let $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p \in \mathbb{R}^n$. Then $\mathbf{u}$ is a *linear combination* of the $\mathbf{v}$; if $\exists\ a_1, a_2, \cdots, a_p \in \mathbb{R}$ such that $\mathbf{u} = a_1\mathbf{v}_1 + \cdots + a_p\mathbf{v}_p$.

3. Subspace

   **Definition 3.** Let $V$ be a set, $V \subset \mathbb{R}^n$. Then $V$ is a subspace if $V$ is closed under addition and scalar multiplication. That is, $V \subset \mathbb{R}^n$ is a subspace if

   (a) $\forall\ \mathbf{v} \in V$ and $\forall\ a \in \mathbb{R}$, $a\mathbf{v} \in V$; and
   (b) $\forall\ \mathbf{v}_1, \mathbf{v}_2 \in V, \mathbf{v}_1 + \mathbf{v}_2 \in V$.

   Alternatively, $V$ is a subspace if $V$ is closed under the operation of taking linear combination (l.c.), i.e., $\forall\ \mathbf{v}_1, \cdots, \mathbf{v}_p \in V$ and for $a_1, \cdots, a_p \in \mathbb{R}, \mathbf{u} = \sum_{i=1}^{p} a_i \mathbf{v}_i \in V$.

   Subspace is a vector space that is a subset of some other (higher-dimension) vector space.

4. Column Space

   The column space contains all linear combinations of the columns of $A$. It is a subspace of $\mathbb{R}^p$ (dimension of the rows) from a column matrix $\mathbf{X}$ of $n \times p$ dimension.

5. Nullspace

   **Definition 4.** The nullspace of a matrix consists of all vectors $\mathbf{x}$ such that $\mathbf{A}\mathbf{x} = \mathbf{0}$. It is denoted by $N(\mathbf{A})$. It is a subspace of $\mathbf{R}^n$ (dimension of the columns), just as the column space was a subspace of $\mathbb{R}^m$.

6. Linear Independence

   Let $\mathbf{v}_1, \cdots, \mathbf{v}_p \in V$ where $V \subset \mathbb{R}^n$ is a subspace. The vectors $\mathbf{v}_i$s, $i = 1, \cdots, p$ are said to be linearly independent if and only if the solution to the linear combinations of $\mathbf{v}_i$s, $a_1\mathbf{v}_1 + \cdots + a_p\mathbf{v}_p = \mathbf{0}$, is $a_i$s $= 0, \forall\ i$.

7. Span of a subspace

   If the vector space $V$ consists of all linear combinations of the vectors $\mathbf{v}_1, \cdots, \mathbf{v}_p$, then these vectors span $V$.

8. Basis

   Let $\mathbf{x}_1, \cdots, \mathbf{x}_p \in V$ where $V \subset \mathbb{R}^n$ is a subspace. If $\mathbf{x}_1, \cdots, \mathbf{x}_p$ spans the subspace $V$, denoted by $\mathcal{L}(\mathbf{x}_1, \cdots, \mathbf{x}_p)$, and that they are linearly independent, then $\mathbf{x}_i$s are the basis vectors and $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_p]$ say, is the basis matrix of $V$.

9. Orthonormal Basis

   A basis with the vectors satisfying $\|\mathbf{v}_i\| = 1$ and $\langle \mathbf{v}_i, \mathbf{v_j} \rangle = 0, i \neq j$. That is, the basis vectors have length one and are orthogonal to each other.

10. Rank

    Let $\mathbf{X}$ be an $n \times p$ matrix. The rank of a matrix is the number of linear independent column or row vectors of $\mathbf{X}$. Specifically,

    - If $n > p$, then the maximum rank of $X$ is $p$.
    - If $n < p$, then the maximum rank of $X$ is $n$.

11. Results from Linear Algebra

    (a) Every subspace has a basis.
    (b) Every basis for a subspace $V$ of $\mathbb{R}^n$ has the same number of elements. This number is called the *dimension* of $V$.
    (c) $\mathbf{v}_1, \cdots, \mathbf{v}_p$ form a basis of $V$, then every vector in $V$ can be written exactly one way as a linear combination of the $\mathbf{v}_i$ (which are the basis vectors).
    (d) If $V$ has dimension $p$, $\mathbf{v}_1, \cdots, \mathbf{v}_p \in V$ and the $\mathbf{v}_i$ are linearly independent, then the $\mathbf{v}_i$ form a basis for $V$.

(e) If $\mathbf{v}_1, \cdots, \mathbf{v}_k$ are linearly independent vectors in a subspace $V$ of dimension $p$, then $p \geq k$.
Note: This result implies that if $k^* > p$, then the $\mathbf{v}_1, \cdots, \mathbf{v}_k$ are linearly dependent.

(f) Let $\mathbf{X}$ be an $n \times p$ matrix of rank $r$, and let

$$V = \{\mathbf{v} \in \mathbb{R}^n : \mathbf{v} = \mathbf{X}\mathbf{a} \text{ for some } \mathbf{a} \in \mathbb{R}^p\}$$

$$W = \{\mathbf{v} \in \mathbb{R}^n : \mathbf{X}'\mathbf{v} = \mathbf{0}\}.$$

Then $\dim(V) = r$ and $\dim(W) = n - r$.

(g) If $W \subset V$ and $\dim(W) = \dim(V)$, then $V = W$.

12.

**Lemma 1.** *Let $\mathbf{X}$ be a basis matrix for the $p$-dimensional subspace $V \subset \mathbb{R}^n$.*

*(a) $\mathbf{X}$ is an $n \times p$ matrix of rank $p$ and $\mathbf{X}'\mathbf{X}$ is invertible.*

*(b) The vector $\mathbf{v} \in V$ iff $\mathbf{v} = \mathbf{X}\mathbf{b}$ for some $\mathbf{b} \in \mathbb{R}^p$. The vector $\mathbf{b}$ is unique.*

13. Vectors Orthogonal To Subspace, and Orthogonal Subspaces

**Definition 5.** Let $\mathbf{u} \in \mathbb{R}^n$ and let $V \subset \mathbb{R}^n$ be a subspace. Then $\mathbf{u}$ is orthogonal to $V$, written $\mathbf{u} \perp V$, if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$, $\forall\ \mathbf{v} \in V$. Let $U$ and $V$ be subspace of $\mathbb{R}^n$. Then $U \perp V$ if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$, $\forall\ \mathbf{u} \in U$, $\forall\ \mathbf{v} \in V$.

14. Subspace Orthogonal Complement

**Definition 6.** Let $V$ be a subspace of $\mathbb{R}^n$. The *orthogonal complement* of $V$ written $V^\perp$, is the set of all vectors orthogonal to $V$. In symbols,

$$V^\perp = \{\mathbf{u} \in \mathbb{R}^n : \langle \mathbf{u}, \mathbf{v} \rangle = 0, \mathbf{v} \in V\}.$$

15. Quotient Space $(\cdot \mod \cdot)$

**Definition 7.** Let $W \subset V$ be a subspace. Then $V \bmod W$, written $V|W$, is the set of all vectors in $V$ that are orthogonal to $W$.

$$V|W = V \cap W^\perp.$$

**Lemma 2.** *$V^\perp$ and $V|W$ are subspaces.*

16. Projection of a vector onto another vector

**Definition 8.** The *projection* of a vector $\mathbf{y}$ on a vector $\mathbf{x}$ is the vector $\mathbf{v}$ such that

(a) $\mathbf{v} = b\mathbf{x}$ for some constant $b$

(b) $\mathbf{y} - \mathbf{v} \perp \mathbf{x}$.

**Remark**: Note that from (2) above,

$$\begin{aligned}
&\langle \mathbf{y} - \mathbf{v}, \mathbf{x} \rangle = 0 \Rightarrow \langle \mathbf{y}, \mathbf{x} \rangle - \langle \mathbf{v}, \mathbf{x} \rangle = 0 \\
\Rightarrow &\langle \mathbf{y}, \mathbf{x} \rangle = \langle \mathbf{v}, \mathbf{x} \rangle \Rightarrow \langle \mathbf{y}, \mathbf{x} \rangle = \langle b\mathbf{x}, \mathbf{x} \rangle \\
\Rightarrow &\langle \mathbf{y}, \mathbf{x} \rangle = b\langle \mathbf{x}, \mathbf{x} \rangle \Rightarrow b = \frac{\langle \mathbf{y}, \mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle} = \frac{\langle \mathbf{y}, \mathbf{x} \rangle}{\|\mathbf{x}\|^2}
\end{aligned}$$

Therefore by first result above, the projection of the vector $\mathbf{y}$ on the vector $\mathbf{x}$ is

$$\mathbf{v} = \begin{cases} \dfrac{\langle \mathbf{y}, \mathbf{x} \rangle}{\|\mathbf{x}\|^2}\mathbf{x}, & \mathbf{x} \neq \mathbf{0} \\ \mathbf{0}, & \mathbf{x} = \mathbf{0} \end{cases}$$

17. Projection of a vector onto a subspace

**Definition 9.** Let $\mathbf{y} \in \mathbb{R}^n$ and $V$ be a subspace of $\mathbb{R}^n$. A projection of $\mathbf{y}$ onto $V$ is a vector $\mathbf{v}$ such that

(a) $\mathbf{v} \in V$

(b) $\mathbf{y} - \mathbf{v} \in V^\perp$

Equivalently, $\mathbf{v}$ is a projection of $\mathbf{y}$ onto $V$ if

(a) $\mathbf{v} \in V$

(b) $(\mathbf{y} - \mathbf{v}) \perp \mathbf{v}$

18. Theorem for Projection

**Theorem 1.** *Let $V$ be a subspace of $\mathbb{R}^n$ and let $\mathbf{y} \in \mathbb{R}^n$.*

*(a) There exists a projection of $\mathbf{y}$ onto $V$.*

*(b) The projection is unique.*

*(c) If $\mathbf{X}$ is a basis matrix for $V$, then the projection of $\mathbf{y}$ onto $V$ is given by $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$*

**Remarks**:

(a) The projection $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is linear in $\mathbf{y}$, since it is of the form $\mathbf{A}\mathbf{y}$. Hence we have shown that projections exits, are unique and are linear functions.

(b) $P_V\mathbf{y}$ is defined to be the projection of $\mathbf{y}$ on the subspace $V$. Hence if $\mathbf{X}$ is the basis matrix for $V$,

$$P_V = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'.$$

$P_V$ is the linear function that assigns to each $\mathbf{y}$ its projection onto $V$.

19. Some Elementary Properties of $P_V$

**Theorem 2** (1.4)**.**

a. ***(Pythogorean Theorem)*** *If $\mathbf{v} \in V$ then $\|\mathbf{y} - \mathbf{v}\|^2 = \|\mathbf{y} - P_V\mathbf{y}\|^2 + \|P_V\mathbf{y} - \mathbf{v}\|^2$*

b. *$\|\mathbf{y} - P_V\mathbf{y}\|^2 \leq \|\mathbf{y} - \mathbf{v}\|^2$ for all $v \in V$ with equality if and only if $\mathbf{v} = P_V\mathbf{y}$.*

*c.* $P_V\mathbf{y} = \mathbf{y}$ *if and only if* $\mathbf{y} \in V$*; and* $P_V\mathbf{y} = \mathbf{0}$ *if and only if* $\mathbf{y} \in V^\perp$

*d.* $P_{V^\perp}\mathbf{y} = \mathbf{y} - P_V\mathbf{y}$*. So that* $\|P_{V^\perp}\mathbf{y}\|^2 = \|\mathbf{y}\|^2 - \|P_V\mathbf{y}\|^2$

*e. Let* $W \subset V$ *be a subspace. Then* $P_W\mathbf{y} = P_W(P_V\mathbf{y}) = P_V(P_W\mathbf{y})$

*f. Let* $W \subset V$ *be a subspace. Then* $P_{V|W}\mathbf{y} = P_V\mathbf{y} - P_W\mathbf{y}$ *and* $\|P_{V|W}\mathbf{y}\|^2 = \|P_V\mathbf{y}\|^2 - \|P_W\mathbf{y}\|^2$

*g. If* $V \perp W$*, then* $P_V(P_W\mathbf{y}) = P_VP_W\mathbf{y} = \mathbf{0}$*. So that* $P_VP_W = \mathbf{0}$*.*

*h.* $P_V^2 = P_V$*, and* $P_V' = P_V$*.*

*i. If* $\dim(V) = p$*, then* $\mathrm{tr}(P_V) = p$*.*

20. Vector Projection on Subspace with Orthogonal Basis

**Theorem 3** ((1.5))**.** *Let* $\mathbf{v}_1, \cdots, \mathbf{v}_p$ *be an orthogonal basis for* $V \subset \mathbb{R}^n$*. Then*

$$P_V\mathbf{y} = \sum_{i=1}^{p} P_{\mathbf{v}_i}\mathbf{y}.$$

21. Gram-Schmidt Orthogonalization

**Definition 10.** Let $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_p$ be a basis for a subspace $V \subset \mathbb{R}^n$ of dimension $p$. For $1 \le i \le p$, let $V_i = \mathcal{L}(\mathbf{x}_1, \cdots, \mathbf{x}_i)$ so that $V_1 \subset V_2 \subset \cdots \subset V_p$ are properly nested spaces. Let

$$\begin{aligned} \mathbf{v}_1 &= \mathbf{x}_1; \\ \mathbf{v}_2 &= \mathbf{x}_2 - P_{V_1}\mathbf{x}_2; \\ \mathbf{v}_3 &= \mathbf{x}_3 - P_{V_2}\mathbf{x}_3; \\ &\vdots \\ \mathbf{v}_{i+1} &= \mathbf{x}_{i+1} - P_{V_i}\mathbf{x}_{i+1} \end{aligned}$$

The $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p$ constitute the orthogonal basis.

22. Mutually Orthogonal Nonzero vectors

**Lemma 3.** *Let* $\mathbf{x}_1, \cdots, \mathbf{x}_n$ *be mutually orthogonal nonzero vectors. Then the* $\mathbf{x}_i$ *are linearly independent.*

**Remarks**:

(a) Thus the vectors obtained by Gram-Schmidth orthogonalization constitute an orthogonal basis.

(b) The basis $\{\mathbf{v}_1, \mathbf{v}_1, \cdots, \mathbf{v}_p\}$ can be made into an orthonormal basis by dividing each vector by its length. Let

$$\mathbf{v}_i^* = \frac{\mathbf{v}_i}{\|\mathbf{v}_i\|}, \quad i = 1, \cdots, p$$

$\{\mathbf{v}_1^*, \mathbf{v}_2^*, \cdots, \mathbf{v}_p^*\}$ constitute an orthonormal basis.

(c) One of the most important aspects of the $P_V$ matrix is that it does not depend on the particular basis matrix chosen. This follows from the uniqueness of the projection. Therefore, it can often be assumed that the basis matrix is orthogonal.

23. Orthonormal Basis Matrix

**Theorem 4** (1.7)**.** *Let* $\mathbf{X}$ *be an orthonormal basis matrix for the subspace* $V$*. Then*

*(a)* $\mathbf{X}'\mathbf{X} = \mathbf{I}$

*(b)* $P_V\mathbf{y} = \mathbf{X}\mathbf{X}'\mathbf{y}$

*(c)* $\|P_V\mathbf{y}\|^2 = \|\mathbf{X}'\mathbf{y}\|^2$

24. Dimension of Quotient Spaces and Orthogonal Complement

**Lemma 4** (1.8)**.** *Let* $W \subset V \subset \mathbb{R}^n$*. Then*

*(a)* $\dim V^\perp = n - \dim(V)$

*(b)* $\dim(V|W) = \dim V - \dim W$

**Lemma 5.**

*(a)* $(V^\perp)^\perp = V$

*(b) Let* $W \subset V$*. then* $V|(V|W) = W$

25. Image of a Matrix

**Definition 11.** Let $\mathbf{A}$ be an $n \times n$ matrix. The image of $\mathbf{A}$ is $V = \{\mathbf{v} \in \mathbb{R}^n : \mathbf{v} = \mathbf{A}\mathbf{c} \text{ for some } \mathbf{c} \in \mathbb{R}^n\}$.

**Remarks**:

(a) Stapleton refers to $V$ as the *column space* or *range space* of $\mathbf{A}$.

(b) $V$ is a subspace.

26. Image of a Matrix as a Projection. Similar to Theorem 2 (h).

**Theorem 5.** *Let* $V =$ *image of a* $\mathbf{A}$*.* $\mathbf{A} = P_V$ *if and only if* $\mathbf{A} = \mathbf{A}'$ *(symmetric) and* $\mathbf{A}^2 = \mathbf{A}$ *(idempotent). If* $\mathbf{A} = P_V$*, then* $\dim(V) = rk(\mathbf{A}) = tr(\mathbf{A})$*.*

**Remark**: Note that the theorem above states that a matrix $\mathbf{A}$ is a population matrix if and only if it is symmetric idempotent.

27. Examples of Projections

(a) Let $\mathbf{1}_n$ denote an $n \times 1$ column vector of 1's and let $\mathcal{L}(\mathbf{1}_n)$ be the subspace of interest. Then the basis matrix is given by $\mathbf{1}_n$ and

$$P_{\mathcal{L}(\mathbf{1}_n)} = \mathbf{1}_n(\mathbf{1}_n'\mathbf{1}_n)^{-1}\mathbf{1}_n' = \mathbf{1}_n(n)^{-1}\mathbf{1}_n' = \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'$$

Suppose $\mathbf{y} \in \mathbb{R}^n$. Then $P_{\mathbf{y}} = \frac{1}{n}\mathbf{1}_n \underbrace{\mathbf{1}_n'\mathbf{y}}_{\sum y_i} = \bar{\mathbf{y}}\mathbf{1}_n$.

(b) Consider $P = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'$. This is the projection operator onto the space that is orthogonal to $\mathcal{L}(\mathbf{1}_n)$.

Recall: $P_{V^\perp}\mathbf{y} = \mathbf{y} - P_V\mathbf{y} \Rightarrow P_{V^\perp}\mathbf{y} = (\mathbf{I} - P_V)\mathbf{y}$, true $\forall\ \mathbf{y} \in \mathbb{R}^n$. Implies that $P_{V^\perp} = (\mathbf{I} - P_V)$.

A vector $\mathbf{v}$ is orthogonal to $\mathbf{1}_n$ if $\langle \mathbf{v}, \mathbf{1}_n \rangle = 0 \Rightarrow \mathbf{1}_n'\mathbf{v} = \sum_1^n v_i = 0$. Thus the subspace that is orthogonal to $\mathcal{L}(\mathbf{1}_n)$ is the subspace of column vectors whose components add to zero.

For this example, $P_{V^\perp}\mathbf{y} = (\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n')\mathbf{y} = \mathbf{y} - \bar{y}\mathbf{1}_n =$ vectors of the deviations $(y_i - \bar{y})$. And $\sum_{i=1}^n (y_i - \bar{y}) = 0$.

# 3 Expectations on Random Vectors

1. Mean Vector

**Definition 12.** Let $\mathbf{Y} = [Y_1, \cdots, Y_n]'$. The mean vector $\mathrm{E}\mathbf{Y}$ is given by

$$\mathrm{E}\mathbf{Y} = [\mathrm{E}Y_1, \cdots, \mathrm{E}Y_n]',$$

where each $Y_i, i = \{1, \cdots, n\}$ is a random variable.

2. Covariance Matrix

**Definition 13.** Let $\mathbf{Y} = [Y_1, \cdots, Y_n]'$ with mean vector $\boldsymbol{\mu}$. Then the covariance matrix of $\mathbf{Y}$, denoted by $\mathrm{cov}(\mathbf{Y})$ is the $n \times n$ matrix whose $(i, j)$ component is $\mathrm{cov}(Y_i, Y_j)$. In symbols,

$$\begin{aligned}\mathrm{cov}(\mathbf{Y}) &= [(\mathrm{cov}(Y_i, Y_j))] \\ &= \begin{bmatrix} \mathrm{var}(Y_1) & \mathrm{cov}(Y_1, Y_2) & \cdots & \mathrm{cov}(Y_1, Y_n) \\ \mathrm{cov}(Y_2, Y_1) & \mathrm{var}(Y_2) & \cdots & \mathrm{cov}(Y_2, Y_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(Y_n, Y_1) & \mathrm{cov}(Y_n, Y_2) & \cdots & \mathrm{var}(Y_n) \end{bmatrix}\end{aligned}$$

**Recall**: For a random vector $\mathbf{Y}$ taking values in $\mathbb{R}^n$, with mean vector $\boldsymbol{\mu}$,

$$\begin{aligned}\mathrm{cov}(\mathbf{Y}) &= \mathrm{E}[(\mathbf{Y} - \mathrm{E}\mathbf{Y})(\mathbf{Y} - \mathrm{E}\mathbf{Y})'] \\ &= \mathrm{E}[(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})'].\end{aligned}$$

(a) $\mathrm{cov}(\mathbf{Y} + \mathbf{a}) = \mathrm{cov}(\mathbf{Y})$

(b) $\mathrm{cov}(\mathbf{A}\mathbf{Y}) = \mathbf{A}\mathrm{cov}(\mathbf{Y})\mathbf{A}'$

(c) If $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$ with $\mathrm{E}\mathbf{X} = \boldsymbol{\mu}$, $\mathrm{cov}(\mathbf{X}) = \boldsymbol{\Sigma}$, then $\mathrm{cov}(\mathbf{Y}) = \mathbf{A}\mathrm{cov}(\mathbf{X})\mathbf{A}'$.

3. Quadratic Form

**Definition 14.** Let $\mathbf{A}$ be a symmetric $n \times n$ matrix. A quadratic form is a function $Q(\mathbf{x})$ defined on $\mathbb{R}^n$ with

$$Q(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x}$$

**Remark**: The requirement that $A$ be symmetric is not a restriction on $Q$, since if $A$ is not symmetric, taking $\mathbf{B} = \frac{1}{2}(\mathbf{A} + \mathbf{A}')$ yields a symmetrix matrix. And it can be shown that

$$Q(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\mathbf{B}\mathbf{x}.$$

Try the following example in Python:

```
import numpy as np
# Given A
A = np.array([1,2,3,4]).reshape([2,2])

# Transform A into symmetric matrix B
B = .5 * (A + A.transpose())

# If X is defined as follows:
X = np.array([8, 9]).reshape([2,1])

# Then the following should be equal
X.transpose().dot(A).dot(X)
X.transpose().dot(B).dot(X)
```

**Corollary 1.** *Let* $\mathbf{a} \in \mathbb{R}^n$*. Then* $\mathbf{a}'\mathrm{cov}(\mathbf{X})\mathbf{a} \geq 0$.

4. Expected Value of Quadratic Form

**Theorem 6.** *Let* $\mathbf{x}$ *be a random vector with mean* $\boldsymbol{\mu}$ *and covariance matrix* $\boldsymbol{\Sigma}$*. Let* $Q(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x}$ *where* $\mathbf{A}$ *is a symmetrix* $n \times n$ *matrix. Then*

$$\begin{aligned}\mathrm{E}[Q(\mathbf{x})] &= \mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}) + Q(\boldsymbol{\mu}) \\ &= \mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}.\end{aligned}$$

**Special Cases**:

(a) If $\mathbf{A} = \mathbf{I}_n$, then $Q(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\mathbf{x} = \|\mathbf{x}\|^2$, and $Q(\boldsymbol{\mu}) = \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu} = \boldsymbol{\mu}'\boldsymbol{\mu} = \|\boldsymbol{\mu}\|^2$. So that

$$\mathrm{E}Q(\mathbf{x}) = \mathrm{E}\|\mathbf{x}\|^2 = \mathrm{tr}(\mathbf{I}_n\boldsymbol{\Sigma}) + \|\boldsymbol{\mu}\|^2 = \mathrm{tr}(\boldsymbol{\Sigma}) + \|\boldsymbol{\mu}\|^2.$$

(b) If $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}_n$, then $\mathrm{E}Q(\mathbf{x}) = \mathrm{tr}(\mathbf{A}\sigma^2\mathbf{I}_n) + Q(\boldsymbol{\mu}) = \sigma^2\mathrm{tr}(\mathbf{A}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}$.

(c) If $\boldsymbol{\mu} = \mathbf{0}$, then $\mathrm{E}Q(\mathbf{x}) = \mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \mathbf{0} = \mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma})$.

5. Expected Value of a Linear Model

**Theorem 7.** *Let* $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$*, where* $\mathbf{Y} \in \mathbb{R}^m$*,* $\mathbf{A}$ *is* $m \times n$*,* $\mathbf{X} \in \mathbb{R}^n$*,* $b \in \mathbb{R}^m$*. Then* $\mathrm{E}\mathbf{Y} = \mathbf{A}\mathrm{E}\mathbf{X} + \mathbf{b}$.

**Remark**: The theorem above states that the expected value of a linear function is same linear function of the expected values. Recal that the inner product is a linear function. Hence,

$$\mathrm{E}\langle \mathbf{a}, \mathbf{Y} \rangle = \langle \mathbf{a}, \mathrm{E}\mathbf{Y} \rangle.$$

6. Nonnegative Definite

**Definition 15.** A symmetrix matrix $\mathbf{U}$ is nonnegative definite written $\mathbf{U} \geq 0$, if $\mathbf{a}'\mathbf{U}\mathbf{a} \geq 0 \ \forall \ \mathbf{a} \in \mathbb{R}^n$. If $\mathbf{a}'\mathbf{U}\mathbf{a} > 0 \ \forall \ \mathbf{a} \neq \mathbf{0}$, $\mathbf{U}$ is positive definite, written $\mathbf{U} > 0$.

**Remark**: The Corollary 1 therefore implies that covariance matrices are nonnegative definite.

7. Positive Definite Covariance

**Lemma 6.** *Let* $\mathrm{cov}(\mathbf{Y}) \geq 0$. *If* $\mathbf{Y}$ *has continuous density function then* $\mathrm{cov}(\mathbf{Y}) > 0$.

8. Joint Moment Generating Function

**Definition 16.** Let $\mathbf{t} = [t_1, \cdots, t_n]'$ be a vector and let $\mathbf{Y} = [Y_1, \cdots, Y_n]'$ be a random vector. Then the joint moment generating function of $\mathbf{Y}$ is defined by

$$\begin{aligned} M_{\mathbf{Y}}(\mathbf{t}) &= \mathrm{E}[\exp(\mathbf{Y}'\mathbf{t})] \\ &= \mathrm{E}\left[\exp\left(\sum_i y_i t_i\right)\right]. \end{aligned}$$

9. Identically Distributed Random Vectors

**Lemma 7.** *If* $\mathbf{X}$ *and* $\mathbf{Y}$ *are random vectors with the same joint moment generating function in some open sets containing the origin, then they have the same distribution.*

10. More Results for Moment Generating Function

**Lemma 8.**

*(a) Let* $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$, *then* $M_{\mathbf{Y}}(\mathbf{t}) = \exp[\mathbf{b}'\mathbf{t}]M_{\mathbf{X}}(\mathbf{A}'\mathbf{Y})$

*(b) Let* $\mathbf{Y} = c\mathbf{Z}, c \in \mathbb{R}$, *then* $M_{\mathbf{Y}}(\mathbf{t}) = M_{\mathbf{Z}}(c\mathbf{t})$.

*(c) Let* $\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix}, \mathbf{t} = \begin{pmatrix} \mathbf{t}_1 \\ \mathbf{t}_2 \end{pmatrix}$, *then*

$$M_{\mathbf{Y}_1}(\mathbf{t}_1) = M_{\mathbf{Y}}\begin{pmatrix} \mathbf{t}_1 \\ \mathbf{0} \end{pmatrix}.$$

*(d)* $\mathbf{Y}_1$ *and* $\mathbf{Y}_2$ *are independent if and only if*

$$M_{\mathbf{Y}}(\mathbf{t}) = M_{\mathbf{Y}_1}\begin{pmatrix} \mathbf{t}_1 \\ \mathbf{0} \end{pmatrix} M_{\mathbf{Y}_2}\begin{pmatrix} \mathbf{0} \\ \mathbf{t}_2 \end{pmatrix}$$

11. More Preliminaries on NND and PD Matrices

**Lemma 9.**

*(a) If* $\mathbf{U} \geq 0$, *then* $\exists$ *a unique* $\mathbf{V} \geq 0$, *such that* $\mathbf{U} = \mathbf{V}^2$. *If* $\mathbf{U} > 0$, *then* $\mathbf{V} > 0$.

*(b) If* $\mathbf{U} > 0$, *then* $\mathbf{U}$ *is invertible and* $\mathbf{U}^{-1} > 0$.

*(c) If* $\mathbf{U} \geq 0$ *then* $\mathbf{U}$ *is invertible if and only if* $\mathbf{U} > 0$.

*(d) If* $\mathbf{U} \geq 0$ *is a* $p \times p$ *matrix of rank* $r$, *then* $\exists$ *a* $p \times r$ *matrix* $C$ *of rank* $r$ *such that* $\mathbf{U} = \mathbf{C}\mathbf{C}'$.

*(e) Let* $\mathbf{X}$ *be an* $n \times p$ *matrix of rank* $r$, *then* $\mathbf{X}'\mathbf{X} \geq 0$. *If* $r = p$, *then* $\mathbf{X}'\mathbf{X} > 0$ *and hence* $\mathbf{X}'\mathbf{X}$ *is invertible.*

12. Square Root of a Matrix

**Definition 17.** Let $\mathbf{A} \geq 0$. The square root of $\mathbf{A}$ is the unique matrix $\mathbf{B} \geq 0$ such that $\mathbf{A} = \mathbf{B}^2$. Notation for square root of $\mathbf{A}$: $\mathbf{A}^{\frac{1}{2}}$.

**Some results**: If $\mathbf{A} > 0$, then

(a) $\mathbf{A}^{-1} > 0, \mathbf{A}^{\frac{1}{2}} > 0$;

(b) $(\mathbf{A}^{-1})^{\frac{1}{2}} = (\mathbf{A}^{\frac{1}{2}})^{-1}$, we define $\mathbf{A}^{-\frac{1}{2}} = (\mathbf{A}^{\frac{1}{2}})^{-1}$;

(c) $\mathbf{A}^{-\frac{1}{2}}\mathbf{A}\mathbf{A}^{-\frac{1}{2}} = \mathbf{I}$, $(\mathbf{A}^{-\frac{1}{2}})' = \mathbf{A}^{-\frac{1}{2}}$, $(\mathbf{A}^{\frac{1}{2}})' = \mathbf{A}^{\frac{1}{2}}$.

13. Loss Function and Mahalanobis Distance

Let $\mathbf{d} \in \mathbb{R}^n$, the loss function we often use for such problems is

$$L(\mathbf{d}; (\boldsymbol{\mu}, \boldsymbol{\Sigma})) = (\mathbf{d} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{d} - \boldsymbol{\mu}).$$

The square root of $L$ is called the mahalanobis distance between $\mathbf{d}$ and $\boldsymbol{\mu}$. $L$ is called the mahalanobis distance loss function.

# 4 Multivariate Normal Distribution

1. The Multivariate Normal Distribution

We define the normal dsitribution by its moment generating function rather than by its density function so as to allow for the possibility of multivariate normal distributions whose covariance matrix is not necessarily positive definite.

**Definition 18.** Let $\boldsymbol{\mu}$ be an $n \times 1$ vector and let $\boldsymbol{\Sigma}$ be a nonnegative definite matrix. We say that the $n$-dimensional random vector $\mathbf{Y}$ has an $n$-dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ if $\mathbf{Y}$ has mgf

$$M_{\mathbf{Y}}(t) = \exp\left[\boldsymbol{\mu}'\mathbf{t} + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}\right].$$

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

2. Distribution of Transformed Random Vector

**Theorem 8.**

*(a) If* $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, *then* $\mathrm{E}\mathbf{Y} = \boldsymbol{\mu}$ *and* $\mathrm{cov}(\mathbf{Y}) = \boldsymbol{\Sigma}$

*(b) If* $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *and* $c \in \mathbb{R}^n$, *then* $c\mathbf{Y} \sim N_n(c\boldsymbol{\mu}, c^2\boldsymbol{\Sigma})$

*(c) Let* $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, *if* $\mathbf{A}$ *is a* $p \times n$, $\mathbf{b}$ *is a* $p \times 1$, *then* $\mathbf{A}\mathbf{Y} + \mathbf{b} \sim N_p(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$

(d) *Let $\boldsymbol{\mu}$ be any $n \times 1$ vector, and let $\boldsymbol{\Sigma}$ be any $n \times n$ nonnegative definite matrix. Then $\exists\, \mathbf{Y}$ such that $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.*

**Remark** We've shown that the formula of normal distribution is preserved under linear operations on the random vectors. We now show that this is preserved under taking marginal and conditional distributions.

3. Marginal of Multivariate Normal Distribution

**Theorem 9.** *Suppose that $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let*

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$$

*where $\mathbf{Y}_1$ and $\boldsymbol{\mu}_1$ are $p \times 1$ and $\boldsymbol{\Sigma}_{11}$ is $p \times p$.*

(a) $\mathbf{Y}_1 \sim N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}), \mathbf{Y}_2 \sim N_{n-p}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$

(b) *$\mathbf{Y}_1$ and $\mathbf{Y}_2$ are independent if and only if $\boldsymbol{\Sigma}_{12} = \mathbf{0}$.*

(c) *If $\boldsymbol{\Sigma}_{22} > 0$, then the conditional distribution of $\mathbf{Y}_1$ given $\mathbf{Y}_2$ is*

$$\mathbf{Y}_1 | \mathbf{Y}_2 \sim N_p(\boldsymbol{\mu} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{Y}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}).$$

**Remark**: If $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \boldsymbol{\Sigma} > 0$, we say that $\mathbf{Y}$ has a nonsingular multivariate normal distribution. If $\boldsymbol{\Sigma}$ is not positive definite we say that $\mathbf{Y}$ has a singular multivariate normal distribution.

**Example 1.** Let $Y \sim N(\mu, \Sigma)$, where $\Sigma \in \mathbb{R}$. Then $Y$ is singular univariate normal distribution if $\Sigma = 0$. Hence the MGF of Y is, $M_Y(t) = \exp(\mu t)$, which implies that $\mathrm{P}(Y = \mu) = 1$, that is $Y$ is degenerate at $\mu$. So the only singular normal distribution is a discrete distribution that degenerate at a point.

4. Noncentral Chi-Square Distribution

**Definition 19.** Let $X_1, \cdots, X_n$ be independent normally distributed random variables with mean $\mu_i \ \forall\, i \in \{1, \cdots, n\}$ and variance 1. Then $\sum_{i=1}^n X_i^2$ has a noncentral chi-square distribution with $n$ degrees of freedom and noncentrality parameter $\delta = \sum_{i=1}^n \mu_i^2$.

Notation: $\chi_n^2(\delta)$

5. Central Chi-Square Density

**Definition 20.** The density of a central $\chi_m^2$ distribution is given by

$$f(y; m) = \frac{y^{m/2-1} \exp[-y/2]}{\Gamma\left(\frac{m}{2}\right) \times 2^{m/2}}, \quad y > 0.$$

**Lemma 10.**

(a) *If $Z$ is a $N_1(0,1)$ random variable, then $Z^2 \sim \chi_1^2(0)$; that is the square of a standard normal distribution is a chi squared random variable.*

(b) *If $X_1, \cdots, X_n$ are independent and $X_i \sim \chi_{p_i}^2$, then $X_1 + \cdots + X_n \sim \chi_{p1+\cdots+p_n}^2$; that is, indepedent chi squared variables add to a chi squared variable, and the degrees of freedom also add.*

6. Noncentral $\chi^2$-Square Density

The noncentral chi-square density is a Poisson mixture of central chi-square densities as given below:

$$f(y; m, \delta) = \sum_{k=0}^{\infty} p(k; \delta) \times f(y; m + 2k)$$

where $y > 0$ and $p(k; \delta) = \frac{\exp\left[\frac{\delta}{2}\right](\delta/2)^k}{k!}$.

Thus the noncentral $\chi^2$ distribution is a weighted average of $\chi^2$ central distributions with Poisson weights.

**Lemma 11.** *If $Y|K \sim \chi_{n+2K}^2(0)$, and $K$ has poisson distribution with parameter $\delta/2$. Then $Y \sim \chi_n^2(\delta)$.*

7. Noncentral $t$-distribution

**Definition 21.** Let $X$ and $Y$ be independent such that $X \sim N_1(\mu, 1)$ and $Y \sim \chi_n^2(0)$. Then

$$t = \frac{X}{\sqrt{\frac{Y}{n}}},$$

is said to have a noncentral $t$ distribution with $n$ degrees of freedom and noncentrality parameter $\mu$.

Notation: $t_n(\mu)$

**Remark**: If $\mu = 0$, then we have a central distribution.

8. Noncentral $F$ Distribution

**Definition 22.** Let $X$ and $Y$ be independent variables such that $X \sim \chi_m^2(\delta)$ and $Y \sim \chi_n^2(0)$. Then,

$$F = \frac{X/m}{Y/n}$$

Notation: $F_{m,n}(\delta)$

**Remark**: If $\delta = 0$, then we have a central $F$ distribution.

9. More on Multivariate Normal Distribution

**Lemma 12.** *Let $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2\mathbf{I})$ where $\mathbf{Y} = [Y_1, \cdots, Y_n]'$ and $\boldsymbol{\mu} = [\mu_1, \cdots, \mu_n]'$ and $\sigma^2 > 0$ is a scalar. Then the $Y_i$ are independent, $Y_i \sim N_1(\mu_i, \sigma^2)$ and*

$$\frac{\|\mathbf{Y}\|^2}{\sigma^2} = \frac{\mathbf{Y}'\mathbf{Y}}{\sigma^2} \sim \chi_n^2\left(\frac{\boldsymbol{\mu}'\boldsymbol{\mu}}{\sigma^2}\right).$$

**Theorem 10.** *Let* $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *with* $\boldsymbol{\Sigma} > 0$. *Then* $\mathbf{Y}$ *has density function*

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right].$$

**Theorem 11.** *Let* $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \boldsymbol{\Sigma} > 0$. *Then*

(a) $\mathbf{Y}'\boldsymbol{\Sigma}^{-1}\mathbf{Y} \sim \chi_n^2(\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})$.

(b) $(\mathbf{Y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu}) \sim \chi_n^2(0)$

10. Spherical Normal Distribution

**Example 2.** Let $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2\mathbf{I})$, then $\mathbf{Y}$ is a spherical normal distribution since it is spherically symmetric about $\boldsymbol{\mu}$. That is, it is unchanged under rotation about $\boldsymbol{\mu}$.

**Theorem 12.** *(2.14) Let* $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2\mathbf{I}), \sigma^2 > 0$. *Then*

(a) $P_V\mathbf{y} \sim N_n(P_V\boldsymbol{\mu}, \sigma^2 P_V)$

(b) *If* $V \perp W$, *then* $P_V\mathbf{Y}$ *and* $P_W\mathbf{Y}$ *are independent.*

**Theorem 13.** *(2.15) Let* $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2\mathbf{I}), \sigma^2 > 0$. *Let* $V$ *be a* $p$*-dimensional subspace of* $\mathbb{R}^n$. *Then*

$$\|P_V\mathbf{Y}\|^2 \sim \sigma^2\chi_p^2\left(\frac{\|P_V\boldsymbol{\mu}\|^2}{\sigma^2}\right)$$

**Corollary 2.** *Let* $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2\mathbf{I}), \sigma^2 > 0$. *If* $\mathbf{A}$ *is idempotent and* $rk(\mathbf{A}) = p$. *Then* $\mathbf{Y}'\mathbf{A}\mathbf{Y} \sim \sigma^2\chi_p^2\left(\frac{\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}}{\sigma^2}\right)$.

**Remarks**: The converse is also true. That is, if $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2\mathbf{I}), \sigma^2 > 0$ and if $\mathbf{A}$ is such that $\mathbf{Y}'\mathbf{A}\mathbf{Y} \sim \sigma^2\chi_p^2\left(\frac{\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}}{\sigma^2}\right)$, then $\mathbf{A}$ is an idempotent matrix of rank $p$.

**Theorem 14.** *Let* $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2\mathbf{I}), \sigma^2 > 0$, *also let* $\mathbf{A}$ *and* $\mathbf{B}$ *be* $p \times n$ *and* $s \times n$ *matrices and let* $\mathbf{C}$ *and* $\mathbf{D}$ *be* $n \times n$ *matrices. Then*

(a) $\mathbf{AY}$ *and* $\mathbf{BY}$ *are independent if and only if* $\mathbf{AB}' = \mathbf{0}$.

(b) *If* $\mathbf{AC} = \mathbf{0}$, *then* $\mathbf{AY}$ *and* $\mathbf{Y}'\mathbf{CY}$ *are independent.*

(c) *If* $\mathbf{CD} = \mathbf{0}$, *then* $\mathbf{Y}'\mathbf{CY}$ *and* $\mathbf{Y}'\mathbf{DY}$ *are independent.*