Week 2: Matrix analysis, Eigenvalues & eigenvectors, and Multivariate Normal distribution.

We denote a set of p random variables $X_i : \Omega \to \mathbb{R}$, $i = 1, \cdots, p$ by the (vector-valued) random variable $\mathbb{X} : \Omega \to \mathbb{R}^p$

$$\mathbb{X} := \begin{bmatrix} X_1 \\ \vdots \\ X_p \end{bmatrix} \qquad \text{"random vector"}$$

The <u>mean</u> or <u>expectation</u> of $\mathbb{X}$ is given by

$$\mathbb{E}[\mathbb{X}] = \begin{bmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \\ \vdots \\ \mathbb{E}[X_p] \end{bmatrix} = \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$$

We always make the assumption that $\mu < \infty$.

A typical measurement involves taking n random samples

$$\{\mathbb{X}_1, \mathbb{X}_2, \ldots, \mathbb{X}_n\}$$

We can express this in matrix form.

$$\mathbb{X} = (\mathbb{X}_1, \mathbb{X}_2, \ldots, \mathbb{X}_n)' = \begin{pmatrix} \mathbb{X}_1 \\ \mathbb{X}_2 \\ \vdots \\ \mathbb{X}_n \end{pmatrix}$$

$$= \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & & & \\ X_{n1} & X_{n2} & & X_{np} \end{pmatrix}$$

Notation: transpose of vector x and matrix A are denoted x' and A', respectively.

A matrix of random variables is called a random matrix.

Expectation of a random matrix $\mathbb{X} = (X_{ij})$ is given by

$$\mathbb{E}[\mathbb{X}] = (\mathbb{E}(X_{ij}))$$

<u>Lemma:</u> Let $\mathbb{X} = (X_{ij})$ and $\mathbb{Y} = (Y_{ij})$ be $n \times p$ random matrices. If A, B and C are constant matrices then:

- $\mathbb{E}[\mathbb{X} + \mathbb{Y}] = \mathbb{E}[\mathbb{X}] + \mathbb{E}[\mathbb{Y}]$. (1)

- $\mathbb{E}[A\mathbb{X}B + C] = A\mathbb{E}[\mathbb{X}]B + C$. (2)

<u>Proof:</u> Choosing an arbitrary (i,j)'th element, LHS of (1) is

$$\mathbb{E}[X_{ij} + Y_{ij}] = \underbrace{\mathbb{E}[X_{ij}] + \mathbb{E}[Y_{ij}]}_{\text{RHS of (1)}}.$$

Since (i,j) was arbitrary it holds for all i, j.

In the same way, (i,j)'th element of LHS of (2) is

$$\mathbb{E}\left[\sum_{k=1}^{n}\sum_{l=1}^{p} a_{ik}X_{kl}b_{lj} + c_{ij}\right] = \underbrace{\sum_{k=1}^{n}\sum_{l=1}^{p} a_{ik}\mathbb{E}[X_{kl}]b_{lj} + c_{ij}}_{= \text{RHS of (2)}}.$$

<u>Note:</u> $A = (a_{ij}) \in \mathbb{R}^{l \times m}$; $B = (b_{jk}) \in \mathbb{R}^{m \times n}$, then.

$$(AB)_{ik} = \sum_{j=1}^{m} a_{ij}b_{jk}. \quad i = 1, \dots, l;\ k = 1, \dots, n.$$

If a $p \times 1$ random vector $\mathbb{X} = (X_1, X_2, \ldots, X_p)'$ has mean $\mu = (\mu_1, \ldots, \mu_p)'$, the <u>covariance matrix of $\mathbb{X}$</u> is defined by

$$\Sigma = Var(\mathbb{X}) = \mathbb{E}\left[(\mathbb{X}-\mu)(\mathbb{X}-\mu)'\right].$$

If a $q \times 1$ random vector $\mathbb{Y} = (Y_1, Y_2, \ldots, Y_q)'$ has mean $\eta = (\eta_1, \eta_2, \ldots, \eta_q)'$, the <u>covariance matrix of $\mathbb{X}$ and $\mathbb{Y}$</u> is defined by

$$Cov(\mathbb{X}, \mathbb{Y}) = \mathbb{E}\left[(\mathbb{X}-\mu)(\mathbb{Y}-\eta)'\right].$$

In particular, $Cov(\mathbb{X}, \mathbb{X}) = Var(\mathbb{X})$.

Elementwise we have. $\Sigma = (\sigma_{ij})$ with

$$\sigma_{ij} = \mathbb{E}\left[(X_i - \mu_i)(X_j - \mu_j)\right]$$
$$\underbrace{= Cov(X_i, X_j)}_{\text{The covariance between } X_i \text{ and } X_j.}$$

$$\sigma_{ii} = \mathbb{E}\left[(X_i - \mu_i)^2\right] = Var(X_i)$$

We write $\sigma_i^2 = \sigma_{ii}$

Theorem: Let Σ be the covariance matrix of a $p\times 1$ random vector $\mathbb{X}$.

(1) Σ is positive semidefinite (nonneg. definite) that is, for any $p\times 1$ fixed vector $x=(x_1,\dots,x_p)'$

$$x'\Sigma x = \sum_{i=1}^{p}\sum_{j=1}^{p}\sigma_{ij}x_i x_j \geq 0.$$

(2) Let B be a $q\times p$ constant matrix and b be a $q\times 1$ constant vector. Then the covariance matrix $\mathbb{Y}=B\mathbb{X}+\mathbb{b}$ is.

$$\operatorname{Var}(\mathbb{Y}) = B\Sigma B'.$$

Proof: (1) For any $x\in\mathbb{R}^p$, we have

$$\begin{aligned}\operatorname{Var}(x'\mathbb{X}) &= \mathbb{E}\left[(x'\mathbb{X}-x'\mu)(x'\mathbb{X}-x'\mu)'\right]\\ &= \mathbb{E}\left[(x'(\mathbb{X}-\mu))(x'(\mathbb{X}-\mu))'\right]\\ &= \mathbb{E}\left[x'(\mathbb{X}-\mu)(\mathbb{X}-\mu)'x\right] = x'\mathbb{E}[(\mathbb{X}-\mu)(\mathbb{X}-\mu)']x\\ &= x'\Sigma x.\end{aligned}$$

and we also know that $\operatorname{Var}(x'\mathbb{X})\geq 0$, so the result follows.

Proof: (2) As $\mathbb{Y} - \mathbb{E}[\mathbb{Y}] = (B\mathbb{X} + b) - (B\mu + b)$
$$= B(\mathbb{X} - \mu)$$
$\mathbb{E}\mathbb{X} = \mu$

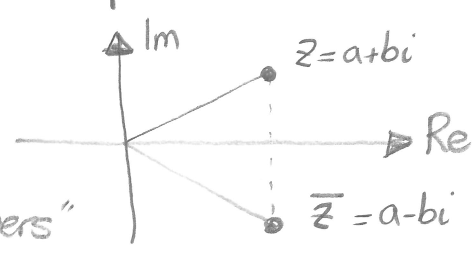
We have
$$\mathrm{Var}(\mathbb{Y}) = \mathbb{E}\left[B(\mathbb{X}-\mu)\left(B(\mathbb{X}-\mu)\right)'\right]$$
$$= \mathbb{E}\left[B(\mathbb{X}-\mu)(\mathbb{X}-\mu)'B'\right]$$
$$= B\,\mathbb{E}\left[(\mathbb{X}-\mu)(\mathbb{X}-\mu)'\right]B'$$
$$= B\Sigma B' \qquad \blacksquare$$

In general, the covariance matrix is positive semidefinite

We call a matrix A positive definite if for all $x \neq 0$,
$$x' A x > 0.$$

Recall: a complex numbers is a number of the form $a+bi$ where i satisfies $i^2 = -1$. We write $\mathrm{Re}(a+bi) = a$
$\mathrm{Im}(a+bi) = b$.

The complex conjugate of a complex number $z = a+bi$ is given by $\bar{z} = a - bi$.



$z \in \mathbb{C}$: "space of complex numbers"

If A is a $m \times n$ matrix with complex entries, then the $n \times m$ matrix A^* is obtained by taking the transpose followed by the complex conjugate of each entry.

$$(A^*)_{ij} = \overline{A_{ji}} \quad \text{or} \quad A^* = (\overline{A})' = \overline{A'}$$

Example: $A = \begin{bmatrix} 1 & -3-i \\ 1+2i & 6i \end{bmatrix}$ $\quad A^* = \begin{bmatrix} 1 & 1-2i \\ -3+i & -6i \end{bmatrix}$

The matrix A^* is called the conjugate transpose of A.

Properties: (1) $(A+B)^* = A^* + B^*$, if A, B have same dims.

(2) $(rA)^* = \overline{r} A^*$, for $r \in \mathbb{C}$ and matrix A.

(3) $(AB)^* = B^* A^*$, $A \in \mathbb{R}^{m \times n}$ $B \in \mathbb{R}^{n \times p}$

(4) $(A^*)^* = A$, $A \in \mathbb{R}^{n \times n}$

(5) If $A \in \mathbb{R}^{n \times n}$, then $\det(A^*) = \overline{(\det A)}$ & $tr(A^*) = \overline{(tr A)}$

Classes of matrices

A Hermitian matrix A is a square matrix that satisfies $A = A^*$.

That means if $A = (a_{ij})$ then $a_{ij} = \overline{a_{ji}}$.

The nice thing about Hermitian matrices is that they behave a bit like real numbers. Arbitrary square matrices behave like complex numbers (ie., they can have some quirky behaviour).

Sometimes I will write $\boxed{A \in M_p}$ to denote that A is a square matrix of size $p \times p$ and if A is Hermitian I will write $\boxed{A \in \mathbb{H}_p}$.

Notice: $\mathbb{H}_p \subseteq M_p$.

We can define the (Frobenius) norm of a matrix $A \in M_p$ as

$$\|A\|_F = \sqrt{\sum_{j=1}^{p} \sum_{k=1}^{p} |a_{jk}|^2} \qquad A \in M_p.$$

Eigenvalues and Eigenvectors

If $A \in M_p$ and if $Ae = \lambda e$ for $e \neq 0$ where $e \in \mathbb{R}^p$ then λ and e are called an eigenvalue and an eigenvector of A.

Example: $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ $e = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ $f = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

$Ae = 3e$ so e is eigenvector, $\lambda = 3$. f is not an eigenvector.

If $A \in M_p$ and $\lambda = (\lambda_1, \dots, \lambda_p)$ "eigenvalues."

$E = (e_1, e_2, \dots, e_p) \in M_p$ "eigenvectors."

Then $AE = \lambda E \iff |A - \lambda I| = 0$ and E satisfies $(A - \lambda I)E = 0$.

If and only if

$|A| = \det(A)$

The expression $p(\lambda) := |A - \lambda I|$ is called the characteristic polynomial for A. The equation $p(\lambda) = 0$ is the characteristic equation for A.

If $B \in M_p$ then the trace is the sum of the diagonal entries, ie.

$$tr\, B := \sum_{k=1}^{p} b_{kk} \quad B \in M_p.$$

Proposition: If $A \in M_p$ then $p(\lambda) = |A - \lambda I|$ is a polynomial in λ of degree p and

Eg. $p = 2$

$$\boxed{P_2(\lambda) = \lambda^2 - tr(A) + \det(A)}$$

$$p(\lambda) = \lambda^p - a_1 \lambda^{p-1} + \dots + (-1)^p a_p$$

$a_1 = Tr(A)$
$a_p = \det(A)$
$a_i =$ "sum of i-rowed diagonal minors of A"

Further: $p(\lambda) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \cdots (\lambda_p - \lambda)$ and $|A| = \prod_{i=1}^{p} \lambda_i$

Example: $B = \begin{pmatrix} 5 & 2 \\ 2 & 5 \end{pmatrix}$

$$|B - \lambda I| = \begin{vmatrix} 5-\lambda & 2 \\ 2 & 5-\lambda \end{vmatrix}$$

$$= 25 - 4 - 10\lambda + \lambda^2$$

$$= \lambda^2 - 10\lambda + 21.$$

A symmetric matrix is a square matrix that is equal to its transpose, ie,

$$A = A'.$$

We denote $A \in \mathbb{S}_p$ and $\mathbb{S}_p \subseteq \mathbb{H}_p$.

Proposition: Let $A \in \mathbb{S}_p$, then

(1) The characteristic roots $\lambda_1, \dots, \lambda_p$ are all real.
$\hookrightarrow$ $\lambda_1, \dots, \lambda_p$ satisfy $p(\lambda_i) = 0$

(2) If λ_i and λ_j are two distinct characteristic roots of A, the corresponding characteristic vectors e_i and e_j are orthogonal.

<u>Proposition</u>: Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ be the characteristic roots of a matrix $A \in \mathcal{S}_p$. Then

(1) $A > 0$ ("positive definite") $\Longleftrightarrow \lambda_i > 0, \; i=1,\dots,p$.

(2) $A \geq 0$ ("positive semidef.") $\Longleftrightarrow \lambda_i \geq 0, \; i=1,\dots,p$.

Given a covariance matrix Σ (positive semidefinite), it follows that the characteristic roots (eigenvalues) are non-negative and these are denoted by

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0.$$

Let γ_i be the characteristic vector (eigenvector) corresponding to λ_i for $i=1,\dots,p$. WLOG, we assume they are <u>orthonormal</u>, ie.,

$$\gamma_i' \gamma_j = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{if } i \neq j. \end{cases}$$

The characteristic roots and vectors satisfy

$$\Sigma \gamma_i = \lambda_i \gamma_i \qquad i=1,\dots,p$$

The relationship $\Sigma \lambda_i = \lambda_i \gamma_i$ can be expressed as

$$\Sigma \Lambda = \Gamma \Lambda \qquad (*)$$

where $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$ and $\Gamma = (\gamma_1, \dots, \gamma_p)$.

We assume that the matrix Γ is normalised as $\Gamma'\Gamma = I_p$, then the equation $(*)$ implies that

$$\begin{aligned} \Sigma &= \Gamma \Lambda \Gamma' \\ &= \Gamma \Lambda^{1/2} \Lambda^{1/2} \Gamma' \\ &= \underbrace{\Gamma \Lambda^{1/2} \Gamma'}\,\underbrace{\Gamma \Lambda^{1/2} \Gamma'} \end{aligned}$$

where $\Lambda^{1/2} = \mathrm{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_p})$.

Define $\Sigma^{1/2} = \Gamma \Lambda^{1/2} \Gamma'$ and $C = \Gamma \Lambda^{1/2}$, then

$$\Sigma = (\Sigma^{1/2})^2 = CC'$$

<u>Note</u>: Here you should be drawing an analogy to variance v. std. dev in univariate case

$$\sigma^2 = \sigma \sigma.$$

The covariance matrix Σ contains the variance of the p variables and the covariances between them. It is desirable to have a measure of "scatter". Two possibilities are:

Generalised variance given in terms of the determinant

(*) $$|\Sigma| = |\Gamma \Lambda \Gamma'| = |\Lambda| = \lambda_1 \cdot \lambda_2 \cdots \lambda_p.$$

Total variance given in terms of trace.

$$\begin{aligned} tr\,\Sigma &= \sigma_{11} + \sigma_{22} + \cdots + \sigma_{pp}. \\ &= tr(\Gamma \Lambda \Gamma') = tr(\Lambda \Gamma' \Gamma) \\ &= tr\,\Lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_n. \end{aligned}$$

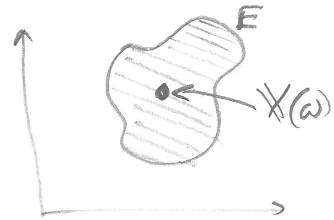
Note: Later we will actually consider the term (*) as

$$av = \frac{1}{p} \log |\Sigma| = \frac{1}{p} \sum_{k=1}^{p} \log(\lambda_k)$$

Characteristic functions.

Let $\mathbb{X}$ be a p-dimensional random vector, then

$$P(\mathbb{X} \in E) = \int_E f(\vec{x})\, d\vec{x}$$



where $d\vec{x} = dx_1 dx_2 \cdots dx_p$. $E \subset \mathbb{R}^p$

The function $f(\vec{x}) = f(x_1, x_2, \cdots, x_p)$ is called the density of $\mathbb{X}$.

The characteristic function of $\mathbb{X}$ is

$$C(\theta) = \mathbb{E}\left[e^{i\theta'\mathbb{X}}\right]$$

$i := \sqrt{-1}$

$\theta := (\theta_1, \theta_2, \cdots \theta_p)'$

$\theta_i \in \mathbb{R}$. $i = 1, \cdots, p$.

Theorem: There exists a one-to-one correspondance between the distribution of $\mathbb{X}$ and its characteristic function.

$$f(\vec{x}) = \frac{1}{(2\pi)^p} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-i\theta'\vec{x}}\, C(\theta)\, d\theta_1 \cdots d\theta_p.$$

You can use the characteristic function to obtain various moments of $\mathbb{X}$.

$$\frac{\partial^m}{\partial\theta_1^{m_1}\partial\theta_2^{m_2}\cdots\partial\theta_p^{m_p}} C(\theta) = \mathbb{E}\left[(iX_1)^{m_1}(iX_2)^{m_2}\cdots(iX_p)^{m_p} e^{i\theta'\mathbb{X}}\right]$$

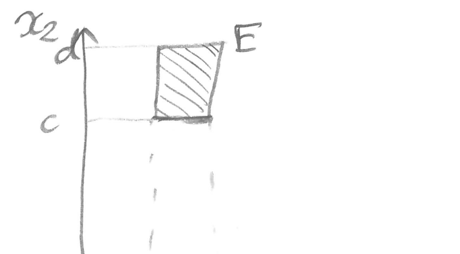
where $m = m_1 + m_2 + \cdots + m_p$.

Taking $\theta = \vec{0}$, we can get the moment $\mathbb{E}\left[X_1^{m_1} X_2^{m_2} \cdots X_p^{m_p}\right]$

Multivariate Normal distribution

The simplest case is the bivariate Normal distribution.

$\mathbb{X} = (X_1, X_2)$



$$P(\mathbb{X} \in E) = P(a < X_1 \leq b, c < X_2 \leq d)$$
$$= \int_a^b \int_c^d f(x_1, x_2)\, dx_1\, dx_2.$$

density given by $f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(\frac{-1}{2(1-\rho^2)} Q\right)$

$$Q(x_1, x_2) = \left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 - 2\rho \frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2$$

$\mathbb{X} \sim$ Bivariate Normal.

$$\mathbb{E}[X_i] = \mu_i \quad Var[X_i] = \sigma_i^2 \qquad i = 1, 2.$$

covariance and correlation between X_1 & X_2 are

$$cov(X_1, X_2) = \rho\sigma_1\sigma_2 \qquad cor(X_1, X_2) = \rho.$$

covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

$$X_1 \sim N(\mu_1, \sigma_1^2) \quad \text{and} \quad X_2 \sim N(\mu_2, \sigma_2^2).$$

Let's construct the <u>higher-dimensional case</u>. $(p > 2)$

<u>Recall:</u>

$$Z \sim N(0,1) \qquad \varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty.$$

if $X \sim N(\mu, \sigma^2)$ then $X = \mu + \sigma Z$. and

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

Take $\mathbb{Z} = (z_1, z_2, \cdots, z_p)'$ $z_i \sim N(0,1)$ iid.

then $\mathbb{Z}$ has density

$$\prod_{i=1}^{p} \frac{1}{\sqrt{2\pi}} e^{-z_i^2/2} = \left(\frac{1}{\sqrt{2\pi}}\right)^p e^{-\vec{z}'\vec{z}/2}$$

$\vec{z} = (z_1, z_2, \cdots, z_p)'$

Now consider $\mathbb{X} = \Sigma^{1/2}\mathbb{Z} + \mu$ $\mu := (\mu_1, \mu_2, \cdots, \mu_p)$

$\Sigma \in \mathbb{S}_p$.

then $\mathbb{E}[\mathbb{X}] = \mu$ $\mathrm{Cov}[\mathbb{X}] = \Sigma$.

<u>Lemma</u>: Let $\vec{f} = \vec{f}(\vec{x}) = (f_1(\vec{x}), f_2(\vec{x}), \cdots, f_p(\vec{x}))'$ be a transformation such that the partial derivatives $\partial f_i / \partial x_j$ exist. Then the determinant of the matrix

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_p} \\ \vdots & \ddots & \\ \frac{\partial f_p}{\partial x_1} & \cdots & \frac{\partial f_p}{\partial x_p} \end{pmatrix}$$

is called the <u>Jacobian determinant</u> and denoted J.

and some useful cases are:

(1) $\vec{y} = A\vec{x} \Rightarrow J = |A|$ $\quad$ $y \in \mathbb{R}^p$, $x \in \mathbb{R}^p$

(2) $Y = AX \Rightarrow J = |A|^q$ $\quad$ $X \in \mathbb{R}^{p\times q}$, $Y \in \mathbb{R}^{p\times q}$

$A \in \mathbb{M}_p \quad B \in \mathbb{M}_q$

(3) $Y = XB \Rightarrow J = |B|^p$

(4) $Y = AXB \Rightarrow J = |A|^q |B|^p$

$\sim$

Now, returning to multivariate Normals, the Jacobian for the transformation from $\mathbb{X}$ to $\mathbb{Z}$ is

$$\Sigma^{-1/2}\mathbb{X} = \Sigma^{-1/2}\Sigma^{1/2}\mathbb{Z} + \Sigma^{-1/2}\mu .$$

as μ doesn't depend on $\mathbb{X}$,

$$\mathbb{Z} = \Sigma^{-1/2}\mathbb{X} - \Sigma^{-1/2}\mu .$$

and $$J = |\Sigma^{-1/2}| = |\Sigma|^{-1/2} .$$

<u>Theorem</u>: If $\mathbb{X}$ is random vector with density $\varphi(\vec{x})$ then the density of $\mathbb{Y} = g(\mathbb{X})$ is given by

$$f(y) = \varphi(g^{-1}(y))\,|J| .$$

Therefore, the density of $\mathbb{X}$ is

$$(*) \qquad f(\vec{x}) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp\left(-\tfrac{1}{2}(\vec{x}-\mu)'\Sigma^{-1}(\vec{x}-\mu)\right).$$

We say that $\mathbb{X}$ has <u>p-variate Normal dist</u> if it has density $(*)$.

<u>Theorem</u>: The following are equivalent:

(1) $\mathbb{X} \sim N_p(\mu, \Sigma)$ $\qquad \Sigma \in M_p$ pos. def.

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

<u>Theorem</u>: $\mathbb{X} \sim N_p(\mu, \Sigma)$

(1) $\mathbb{E}[\mathbb{X}] = \mu \quad Var(\mathbb{X}) = \Sigma$.

(2) $C_{\mathbb{X}}(\theta) = \exp\left(i\mu'\theta - \frac{1}{2}\theta'\Sigma\theta\right)$.

(3) $B \in \mathbb{R}^{q\times p}$, $\text{rank}(B) = q$, $b \in \mathbb{R}^q$
$\mathbb{Y} = B\mathbb{X} + b \sim N_q(B\mu + b, B\Sigma B')$