Random Vectors, Multivariate Normality and Random Samples

1 Review of expectation, variance and covariance

Let's focus on the case of continuous random variables. Let $f_X(x)$ be the pdf of the random variable X. Its expectation is defined as

$$\mathbb{E}[X] := \mu_X := \int_{-\infty}^{\infty} x f_X(x) dx,$$

and its variance is defined as

$$\mathrm{Var}(X) := \mathbb{E}\left[(X-\mu_X)^2\right] = \int_{-\infty}^{\infty} (x-\mu_X)^2 f_X(x) dx.$$

Let X and Y be two jointly distributed random variables with joint pdf $f_{X,Y}(x,y)$. Their covariance is defined as

$$\mathrm{Cov}(X,Y) = \mathbb{E}\left((X-\mathbb{E}(X))(Y-\mathbb{E}(Y))\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y) f_{X,Y}(x,y) dx dy.$$

Properties of expectation, variance, covariance:

1. $\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.
2. Covariances: $\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.
3. $\mathbb{E}(a + \sum_{i=1}^n b_i X_i) = a + \sum_{i=1}^n b_i \mathbb{E}(X_i)$.
4. $\mathrm{Var}(a + bX) = b^2 \mathrm{Var}(X)$.
5. $\mathrm{Var}(X) = \mathrm{Cov}(X,X)$.
6. $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$.
7. $\mathrm{Cov}(a + \sum_{i=1}^n b_i X_i, c + \sum_{j=1}^m d_j Y_j) = \sum_{i=1}^n \sum_{j=1}^m b_i d_j \mathrm{Cov}(X_i, Y_j)$.
8. $\mathrm{Var}(a + \sum_{i=1}^n b_i X_i) = \sum_{i=1}^n b_i^2 \mathrm{Var}(X_i) + 2 \sum\limits_{1\le i<j\le n} b_i b_j \mathrm{Cov}(X_i, X_j)$.
9. If X and Y are independent, then

$$\begin{cases} \mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y) \\ \mathrm{Cov}(X,Y) = 0 \\ \mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) \end{cases}$$

Normal distribution and properties The pdf of $\mathcal{N}(\mu, \sigma^2)$ is

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

In particular, $\mathcal{N}(0,1)$ is referred to as the standard normal distribution. Let $X \sim \mathcal{N}(\mu, \sigma^2)$. We have the following properties:

1. $\mathbb{E}(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$;
2. If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.
3. If $X_1, \ldots, X_n$ are independent and $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, then
$$a + \sum_{i=1}^{n} b_i X_i \sim \mathcal{N}\left(a + \sum_{i=1}^{n} b_i\mu_i, \sum_{i=1}^{n} b_i^2\sigma_i^2\right)$$
4. If $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, then
$$\sum_{i=1}^{n} X_i^2 \sim \chi_n^2.$$

2 Random vectors and matrices

Confidence region for "true" mean $\vec{\mu}$.

A vector

$$\vec{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$$

is referred to as a random vector, if $X_1, \ldots, X_p$ are jointly distributed random variables. Its expectation or **population mean** is defined as

$$\mathbb{E}\,\vec{X} = \begin{bmatrix} \mathbb{E}\,X_1 \\ \mathbb{E}\,X_2 \\ \vdots \\ \mathbb{E}\,X_p \end{bmatrix} := \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} := \vec{\mu}$$

A matrix

$$\boldsymbol{X} = \begin{bmatrix} X_{11} & X_{12} & \ldots & X_{1m} \\ X_{21} & X_{22} & \ldots & X_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ X_{k1} & X_{k2} & \ldots & X_{km} \end{bmatrix}$$

is a random matrix, if $X_{11}, X_{12}, \ldots, X_{km}$ are jointly distributed random variables. Its expectation is defined as

$$\mathbb{E}\,\boldsymbol{X} = \begin{bmatrix} \mathbb{E}\,X_{11} & \mathbb{E}\,X_{12} & \ldots & \mathbb{E}\,X_{1m} \\ \mathbb{E}\,X_{21} & \mathbb{E}\,X_{22} & \ldots & \mathbb{E}\,X_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}\,X_{k1} & \mathbb{E}\,X_{k2} & \ldots & \mathbb{E}\,X_{km} \end{bmatrix}$$

Random vectors $\vec{X}, \vec{Y} \in \mathbb{R}^p$,

$$\mathbb{E}\left(\vec{X}+\vec{Y}\right) = \mathbb{E}\,\vec{X} + \mathbb{E}\,\vec{Y}.$$

Random vector $\vec{X} \in \mathbb{R}^p$, deterministic $\vec{a} \in \mathbb{R}^p$ and $c \in \mathbb{R}$,

$$\mathbb{E}\left(\vec{a}^\top \vec{X} + c\right) = \vec{a}^\top\,\mathbb{E}\,\vec{X} + c.$$

Random vector $\vec{X} \in \mathbb{R}^p$, deterministic $\boldsymbol{C} \in \mathbb{R}^{q\times p}$ and $\vec{d} \in \mathbb{R}^q$,

$$\mathbb{E}(\boldsymbol{C}\vec{X} + \vec{d}) = \boldsymbol{C}\,\mathbb{E}\,\vec{X} + \vec{d}.$$

Random matrices $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{k\times m}$,

$$\mathbb{E}\left(\boldsymbol{X}+\boldsymbol{Y}\right) = \mathbb{E}\,\boldsymbol{X} + \mathbb{E}\,\boldsymbol{Y}$$

Random matrix $\boldsymbol{X} \in \mathbb{R}^{k\times m}$, deterministic $\boldsymbol{A} \in \mathbb{R}^{l\times k}$ and $\boldsymbol{B} \in \mathbb{R}^{m\times n}$,

$$\mathbb{E}\left(\boldsymbol{AXB}\right) = \boldsymbol{A}\left(\mathbb{E}\,\boldsymbol{X}\right)\boldsymbol{B}.$$

3 Population covariance matrix

Population covariance matrix Let $\vec{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} \in \mathbb{R}^p$ be a random vector with population mean $\mathbb{E}\,\vec{X} = \vec{\mu}$. Denote $\mathrm{Cov}(X_j, X_k) = \sigma_{jk}$. Define the population covariance matrix of $\vec{X}$ as

$$\mathrm{Cov}(\vec{X}) = \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{bmatrix}$$

We can derive the following formula:

$$\begin{aligned}
\boldsymbol{\Sigma} &= \begin{bmatrix} \mathrm{Cov}(X_1,X_1) & \mathrm{Cov}(X_1,X_2) & \dots & \mathrm{Cov}(X_1,X_p) \\ \mathrm{Cov}(X_2,X_1) & \mathrm{Cov}(X_2,X_2) & \dots & \mathrm{Cov}(X_2,X_p) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_p,X_1) & \mathrm{Cov}(X_p,X_2) & \dots & \mathrm{Cov}(X_p,X_p) \end{bmatrix} \\
&= \begin{bmatrix} \mathbb{E}\left((X_1-\mu_1)(X_1-\mu_1)\right) & \dots & \mathbb{E}\left((X_1-\mu_1)(X_p-\mu_p)\right) \\ \vdots & \ddots & \vdots \\ \mathbb{E}\left((X_p-\mu_p)(X_1-\mu_1)\right) & \dots & \mathbb{E}\left((X_p-\mu_p)(X_p-\mu_p)\right) \end{bmatrix} \\
&= \mathbb{E}\begin{bmatrix} \left((X_1-\mu_1)(X_1-\mu_1)\right) & \dots & \left((X_1-\mu_1)(X_p-\mu_p)\right) \\ \vdots & \ddots & \vdots \\ \left((X_p-\mu_p)(X_1-\mu_1)\right) & \dots & \left((X_p-\mu_p)(X_p-\mu_p)\right) \end{bmatrix} \\
&= \mathbb{E}\left(\begin{bmatrix} X_1-\mu_1 \\ \vdots \\ X_p-\mu_p \end{bmatrix} \begin{bmatrix} X_1-\mu_1 & \dots & X_p-\mu_p \end{bmatrix}\right) \\
&= \mathbb{E}\left((\vec{X}-\mathbb{E}\,\vec{X})(\vec{X}-\mathbb{E}\,\vec{X})^\top\right).
\end{aligned}$$

Population cross-covariance matrix Let

$$\vec{X} := \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} \quad \text{and} \quad \vec{Y} := \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_q \end{bmatrix}$$

be two jointly distributed random vectors with population means $\vec{\mu}_x$ and $\vec{\mu}_y$ Define the population cross-covariance matrix between $\vec{X}$ and $\vec{Y}$:

$$\operatorname{Cov}(\vec{X}, \vec{Y}) = \begin{bmatrix} \operatorname{Cov}(X_1, Y_1) & \operatorname{Cov}(X_1, Y_2) & \dots & \operatorname{Cov}(X_1, Y_q) \\ \operatorname{Cov}(X_2, Y_1) & \operatorname{Cov}(X_2, Y_2) & \dots & \operatorname{Cov}(X_2, Y_q) \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{Cov}(X_p, Y_1) & \operatorname{Cov}(X_p, Y_2) & \dots & \operatorname{Cov}(X_p, Y_q) \end{bmatrix}$$

We can derive the following formula:

$$\operatorname{Cov}(\vec{X}, \vec{Y}) = \mathbb{E}\left((\vec{X} - \mathbb{E}\,\vec{X})(\vec{Y} - \mathbb{E}\,\vec{Y})^\top\right) \in \mathbb{R}^{p \times q}.(Why?)$$

Properties of population covariance

1. $\operatorname{Cov}(\vec{X}) = \operatorname{Cov}(\vec{X}, \vec{X})$.

2. $\operatorname{Cov}(b\vec{X} + \vec{d}) = b^2 \operatorname{Cov}(\vec{X})$.

3. $\operatorname{Cov}(\vec{X}, \vec{Y}) = \operatorname{Cov}(\vec{Y}, \vec{X})^\top$.

4. $\operatorname{Cov}(\boldsymbol{C}\vec{X} + \vec{d}) = \boldsymbol{C}\operatorname{Cov}(\vec{X})\boldsymbol{C}^\top$.

   Proof. The population mean of $\boldsymbol{C}\vec{X} + \vec{d}$ is

   $$\mathbb{E}\left(\boldsymbol{C}\vec{X} + \vec{d}\right) = \boldsymbol{C}\,\mathbb{E}(\vec{X}) + \vec{d} = \boldsymbol{C}\vec{\mu} + \vec{d}.$$

   The population covariance matrix of $\boldsymbol{C}\vec{X} + \vec{d}$ is

   $$\begin{aligned} \operatorname{Cov}(\boldsymbol{C}\vec{X} + \vec{d}) &= \mathbb{E}\left((\boldsymbol{C}\vec{X} + \vec{d} - (\boldsymbol{C}\vec{\mu} + \vec{d}))(\boldsymbol{C}\vec{X} + \vec{d} - (\boldsymbol{C}\vec{\mu} + \vec{d}))^\top\right) \\ &= \mathbb{E}\,\boldsymbol{C}(\vec{X} - \vec{\mu})(\vec{X} - \vec{\mu})^\top \boldsymbol{C}^\top \\ &= \boldsymbol{C}\left(\mathbb{E}(\vec{X} - \vec{\mu})(\vec{X} - \vec{\mu})^\top\right)\boldsymbol{C}^\top \\ &= \boldsymbol{C}\boldsymbol{\Sigma}_x\boldsymbol{C}^\top \end{aligned}$$

   □

5. $\operatorname{Cov}(\boldsymbol{C}\vec{X}, \boldsymbol{D}\vec{Y}) = \boldsymbol{C}\operatorname{Cov}(\vec{X}, \vec{Y})\boldsymbol{D}^\top$. (Why?)

6. $\operatorname{Cov}\left(\sum_{i=1}^n b_i\vec{X}_i, \sum_{j=1}^m d_j\vec{Y}_j\right) = \sum_{i=1}^n \sum_{j=1}^m b_i d_j \operatorname{Cov}\left(\vec{X}_i, \vec{Y}_j\right)$.

Proof. To show this result, for any $i = 1, \ldots, n$ and $j = 1, \ldots, m$, we denote

$$\vec{X}_i = \begin{bmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{ip} \end{bmatrix} \quad \vec{Y}_j = \begin{bmatrix} Y_{j1} \\ Y_{j2} \\ \vdots \\ Y_{jp} \end{bmatrix}$$

The (ℓk)-entry of the left is

$$\operatorname{Cov}\left(\sum_{i=1}^{n} b_i X_{i\ell}, \sum_{j=1}^{m} d_j Y_{jk}\right).$$

The (ℓk)-entry of the left is

$$\sum_{i=1}^{n}\sum_{j=1}^{m} b_i d_j \operatorname{Cov}\left(X_{i\ell}, Y_{jk}\right).$$

Therefore, the left equals the right. $\square$

7. If $\vec{X} \in \mathbb{R}^p$ and $\vec{Y} \in \mathbb{R}^q$ are independent random vectors, then $\operatorname{Cov}(\vec{X}, \vec{Y}) = \mathbf{0}_{p\times q}$. (Why?)

8. For mutually independent random vectors $\vec{X}_1, \ldots, \vec{X}_n \in \mathbb{R}^p$, we have

$$\operatorname{Cov}(a_1\vec{X}_1 + \ldots + a_n\vec{X}_n + \vec{c}) = a_1^2\operatorname{Cov}(\vec{X}_1) + \ldots + a_n^2\operatorname{Cov}(\vec{X}_n).$$

(Why?)

4 Multivariate normality

Definition 1. *A $p \times 1$ random vector $\vec{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} \sim \mathcal{N}_p(\vec{\mu}, \mathbf{\Sigma})$, if and only if its pdf is*

$$f(\vec{x}) = \frac{1}{(2\pi)^{\frac{p}{2}}|\mathbf{\Sigma}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^\top\mathbf{\Sigma}^{-1}(\vec{x}-\vec{\mu})\right)$$

In this case, $X_1, \ldots, X_p$ are said to jointly normally distributed as $\mathcal{N}(\vec{\mu}, \mathbf{\Sigma})$. We also have $\mathbb{E}[\vec{X}] = \vec{\mu}$ and $\operatorname{Cov}(\vec{X}) = \mathbf{\Sigma}$. Moreover, the distribution $\mathcal{N}(\vec{0}, \boldsymbol{I}_p)$ is referred to as the p-dimensional standard normal distribution.

Property 1: If $\vec{X} \sim \mathcal{N}_p(\vec{\mu}, \mathbf{\Sigma})$, then for $\boldsymbol{C} \in \mathbb{R}^{q\times p}$, $\vec{d} \in \mathbb{R}^q$, then

$$\boldsymbol{C}\vec{X} + \vec{d} \sim \mathcal{N}_q(\boldsymbol{C}\vec{\mu} + \vec{d}, \boldsymbol{C}\mathbf{\Sigma}\boldsymbol{C}^\top).$$

Property 2: If $\vec{X} \sim \mathcal{N}_p(\vec{\mu}, \mathbf{\Sigma})$, then for $\vec{a} \in \mathbb{R}^p$, $b \in \mathbb{R}$, then

$$a_1X_1 + \ldots + a_pX_p + b = \vec{a}^\top\vec{X} + b \sim \mathcal{N}(\vec{a}^\top\vec{\mu} + b, \vec{a}^\top\mathbf{\Sigma}\vec{a}).$$

Property 3: If $\vec{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} \sim \mathcal{N}_p(\vec{\mu}, \boldsymbol{\Sigma})$, then $X_1, X_2, \ldots, X_p$ are mutually independent if and only if $\boldsymbol{\Sigma}$ is a diagonal matrix, i.e., $\text{Cov}(X_j, X_k) = 0$ for any $j \neq k$.

Property 4: Suppose $\boldsymbol{\Sigma}$ is positive definite. If $\vec{X} \sim \mathcal{N}_p(\vec{\mu}, \boldsymbol{\Sigma})$, then

$$\left(\vec{X} - \vec{\mu}\right)^\top \boldsymbol{\Sigma}^{-1} \left(\vec{X} - \vec{\mu}\right) \sim \chi^2_p.$$

Proof. Let $\vec{Z} = \boldsymbol{\Sigma}^{-\frac{1}{2}} \left(\vec{X} - \vec{\mu}\right)$. We have $\mathbb{E}(\vec{Z}) = \boldsymbol{\Sigma}^{-\frac{1}{2}} \left(\vec{\mu} - \vec{\mu}\right) = \vec{0}$ and $\text{Cov}(\vec{Z}) = \boldsymbol{\Sigma}^{-\frac{1}{2}} \text{Cov}(\vec{X}) \boldsymbol{\Sigma}^{-\frac{1}{2}} = \boldsymbol{I}$. This implies that $\vec{Z} \sim \mathcal{N}_p(\vec{0}, \boldsymbol{I})$, i.e., $Z_1, \ldots, Z_p \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$. Then

$$\begin{aligned}
\left(\vec{X} - \vec{\mu}\right)^\top \boldsymbol{\Sigma}^{-1} \left(\vec{X} - \vec{\mu}\right) &= \left(\vec{X} - \vec{\mu}\right)^\top \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\Sigma}^{-\frac{1}{2}} \left(\vec{X} - \vec{\mu}\right) \\
&= \left(\boldsymbol{\Sigma}^{-\frac{1}{2}} \left(\vec{X} - \vec{\mu}\right)\right)^\top \boldsymbol{\Sigma}^{-\frac{1}{2}} \left(\vec{X} - \vec{\mu}\right) \\
&= \vec{Z}^\top \vec{Z} \\
&= Z_1^2 + \ldots + Z_p^2 \\
&\sim \chi^2_p
\end{aligned}$$

□

Property 5: If $\vec{X}_1, \ldots, \vec{X}_n$ are independent multivariate normal random vectors, then $\vec{X}_1 + \ldots + \vec{X}_n$ is a multivariate normal random vector.

Property 6: (The marginal distributions of a multivariate normal distribution are also multivariate normal distributions.) Suppose $\vec{X} \sim \mathcal{N}_p(\vec{\mu}, \boldsymbol{\Sigma})$. Considering the partition

$$\vec{X} = \begin{bmatrix} \vec{X}^{(1)} \\ \hdashline \vec{X}^{(2)} \end{bmatrix},$$

where $\vec{X}^{(1)} \in \mathbb{R}^q$ and $\vec{X}^{(2)} \in \mathbb{R}^{p-q}$. Correspondingly

$$\vec{\mu} = \begin{bmatrix} \vec{\mu}^{(1)} \\ \hdashline \vec{\mu}^{(2)} \end{bmatrix}, \quad \boldsymbol{\Sigma} = \left[\begin{array}{c:c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hdashline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array}\right].$$

Then $\vec{X}^{(1)} \sim \mathcal{N}\left(\vec{\mu}^{(1)}, \boldsymbol{\Sigma}_{11}\right)$ and $\vec{X}^{(2)} \sim \mathcal{N}\left(\vec{\mu}^{(2)}, \boldsymbol{\Sigma}_{22}\right)$. Moreover, for any $\vec{x} \in \mathbb{R}^{p-q}$,

$$\vec{X}^{(1)} | (\vec{X}^{(2)} = \vec{x}) \sim \mathcal{N}_{p_1}\left(\vec{\mu}^{(1)} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\left(\vec{x} - \vec{\mu}^{(2)}\right), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right).$$

5 Random samples

Recall that for an observed (fixed) sample $\vec{x}_1, \ldots, \vec{x}_n$ from a p-variate population, they form a data matrix

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1p} \\ x_{21} & x_{22} & \ldots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{np} \end{bmatrix} = \begin{bmatrix} \vec{x}_1^\top \\ \vec{x}_2^\top \\ \vdots \\ \vec{x}_n^\top \end{bmatrix}.$$

Its sample mean and sample covariance matrix have the fomulas

$$\overline{\vec{x}} := \frac{1}{n}\sum_{i=1}^{n} \vec{x}_i = \frac{1}{n}\boldsymbol{X}^\top \vec{1}_n,$$

and

$$\boldsymbol{S} := \frac{1}{n-1}\sum_{i=1}^{n} \left(\vec{x}_i - \overline{\vec{x}}\right)\left(\vec{x}_i - \overline{\vec{x}}\right)^\top = \frac{1}{n-1}(\boldsymbol{X} - \vec{1}_n\overline{\vec{x}}^\top)^\top(\boldsymbol{X} - \vec{1}_n\overline{\vec{x}}^\top).$$

In contrast, we say $\vec{X}_1, \dots, \vec{X}_n$ is a random sample of the p-variate population $\mathcal{F}$, if they are independent and identically distributed random vectors from $\mathcal{F}$. ***A random sample is used to model unobserved data that are intended to be observed.*** For example, assume the p-dimensional population is represented by a p-dimensional random vector

$$\vec{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_p \end{bmatrix} \sim \mathcal{N}(\vec{\mu}, \boldsymbol{\Sigma}),$$

where X_j represents the j-th variable. A collection of random vectors $\vec{X}_1, \dots, \vec{X}_n \overset{i.i.d.}{\sim} \mathcal{N}(\vec{\mu}, \boldsymbol{\Sigma})$ is referred to as a random sample of $\vec{X}$ with size n. As with the fixed sample, each random vector is denoted as

$$\vec{X}_i = \begin{bmatrix} X_{i1} \\ X_{i1} \\ \vdots \\ X_{ip} \end{bmatrix},$$

and the random data matrix is represented as

$$\boldsymbol{X} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1p} \\ X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix} = \begin{bmatrix} \vec{X}_1^\top \\ \vec{X}_2^\top \\ \vdots \\ \vec{X}_n^\top \end{bmatrix}.$$

We can similarly define the sample mean and sample covariance

$$\overline{\vec{X}} := \frac{1}{n}\sum_{i=1}^{n} \vec{X}_i = \frac{1}{n}\boldsymbol{X}^\top \vec{1}_n.$$

and the sample covariance matrix

$$\boldsymbol{S} := \frac{1}{n-1}\sum_{i=1}^{n} \left(\vec{X}_i - \overline{\vec{X}}\right)\left(\vec{X}_i - \overline{\vec{X}}\right)^\top = \frac{1}{n-1}\left(\boldsymbol{X} - \vec{1}_n\overline{\vec{X}}^\top\right)^\top\left(\boldsymbol{X} - \vec{1}_n\overline{\vec{X}}^\top\right),$$

5.1 Algebraic properties

All the algebraic properties on fixed samples hold the same for random samples. For example, the linear transformation on the variates

$$\underset{(q\times 1)}{\vec{Y}} = \underset{(q\times p)}{\boldsymbol{C}}\underset{(p\times 1)}{\vec{X}} + \underset{(q\times 1)}{\vec{d}}$$

yields the linear transformation on the random sample

$$\vec{Y}_i = \boldsymbol{C}\vec{X}_i + \vec{d}, \quad i = 1, \dots, n.$$

The sample covariance of $\{\vec{Y}_i\}_{i=1}^n$ is then

$$\boldsymbol{S}_Y = \boldsymbol{C}\boldsymbol{S}_X\boldsymbol{C}^\top$$

where $\boldsymbol{S}_X$ is the sample covariance of $\{\vec{X}_i\}_{i=1}^n$.

In particular, under the linear combination

$$Y = \vec{b}^\top \vec{X}_i = b_1 X_{i1} + \ldots b_p \vec{X}_{ip},$$

the sample variance Y is $S_Y^2 = \vec{b}^\top \boldsymbol{S}_X \vec{b}$.

5.2 Probabilistic properties

Let $\vec{X}_1, \ldots, \vec{X}_n$ be a random sample from $\mathcal{N}_p(\vec{\mu}, \boldsymbol{\Sigma})$. We have the following properties on $\boldsymbol{S}$ and $\overline{\vec{X}}$:

1. $\mathbb{E}\left(\overline{\vec{X}}\right) = \vec{\mu}$;

   Proof.
   $$\mathbb{E}\left(\overline{\vec{X}}\right) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n \vec{X}_i\right) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left(\vec{X}_i\right) = \frac{1}{n}(n\vec{\mu}) = \vec{\mu}.$$
   □

2. $\mathrm{Cov}\left(\overline{\vec{X}}\right) = \frac{1}{n}\boldsymbol{\Sigma}$;

   Proof. Since $\vec{X}_1, \ldots, \vec{X}_n$ are independent, we have
   $$\mathrm{Cov}\left(\overline{\vec{X}}\right) = \mathrm{Cov}\left(\frac{1}{n}\sum_{i=1}^n \vec{X}_i\right) = \frac{1}{n^2}\mathrm{Cov}\left(\sum_{i=1}^n \vec{X}_i\right) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Cov}\left(\vec{X}_i\right) = \frac{1}{n^2}\left(n\boldsymbol{\Sigma}\right) = \frac{1}{n}\boldsymbol{\Sigma}.$$
   □

3. $\overline{\vec{X}} \sim \mathcal{N}_p\left(\vec{\mu}, \frac{1}{n}\boldsymbol{\Sigma}\right)$;

   Proof. Since $\overline{\vec{X}} = \sum_{i=1}^n \frac{1}{n}\vec{X}_i$, and $\vec{X}_1, \ldots, \vec{X}_n$ are independent multivariate normal random vectors, we know $\overline{\vec{X}}$ is a multivariate normal random vector. □

4. $n(\overline{\vec{X}} - \vec{\mu})^\top \boldsymbol{\Sigma}^{-1}(\overline{\vec{X}} - \vec{\mu}) \sim \chi_p^2$;

   Proof. This is simply due to $\overline{\vec{X}} \sim \mathcal{N}_p\left(\vec{\mu}, \frac{1}{n}\boldsymbol{\Sigma}\right)$. □

5. $\mathbb{E}(\boldsymbol{S}) = \boldsymbol{\Sigma}$;

Proof. Method 1: To evaluate $\mathbb{E}(\boldsymbol{S})$, we first have

$$
\begin{aligned}
&(n-1)\boldsymbol{S}\\
&=\sum_{i=1}^{n}(\vec{X}_i-\overline{\vec{X}})(\vec{X}_i-\overline{\vec{X}})^\top\\
&=\sum_{i=1}^{n}\left(\left(\vec{X}_i-\vec{\mu}\right)-\left(\overline{\vec{X}}-\vec{\mu}\right)\right)\left(\left(\vec{X}_i-\vec{\mu}\right)-\left(\overline{\vec{X}}-\vec{\mu}\right)\right)^\top\\
&=\sum_{i=1}^{n}\left(\left(\vec{X}_i-\vec{\mu}\right)\left(\vec{X}_i-\vec{\mu}\right)^\top-\left(\overline{\vec{X}}-\vec{\mu}\right)\left(\vec{X}_i-\vec{\mu}\right)^\top-\left(\vec{X}_i-\vec{\mu}\right)\left(\overline{\vec{X}}-\vec{\mu}\right)^\top+\left(\overline{\vec{X}}-\vec{\mu}\right)\left(\overline{\vec{X}}-\vec{\mu}\right)^\top\right)\\
&=\sum_{i=1}^{n}\left(\vec{X}_i-\vec{\mu}\right)\left(\vec{X}_i-\vec{\mu}\right)^\top-\left(\overline{\vec{X}}-\vec{\mu}\right)\left(\sum_{i=1}^{n}\vec{X}_i-n\vec{\mu}\right)^\top-\left(\sum_{i=1}^{n}\vec{X}_i-n\vec{\mu}\right)\left(\overline{\vec{X}}-\vec{\mu}\right)^\top+n(\overline{\vec{X}}-\vec{\mu})(\overline{\vec{X}}-\vec{\mu})^\top\\
&=\sum_{i=1}^{n}\left(\vec{X}_i-\vec{\mu}\right)\left(\vec{X}_i-\vec{\mu}\right)^\top-n(\overline{\vec{X}}-\vec{\mu})(\overline{\vec{X}}-\vec{\mu})^\top
\end{aligned}
$$

This implies that

$$
\begin{aligned}
\mathbb{E}((n-1)\boldsymbol{S})&=\sum_{i=1}^{n}\mathbb{E}\left(\vec{X}_i-\vec{\mu}\right)\left(\vec{X}_i-\vec{\mu}\right)^\top-n\,\mathbb{E}\left(\overline{\vec{X}}-\vec{\mu}\right)\left(\overline{\vec{X}}-\vec{\mu}\right)^\top\\
&=\sum_{i=1}^{n}\mathrm{Cov}(\vec{X}_i)-n\frac{1}{n}\mathrm{Cov}(\overline{\vec{X}})\\
&=\sum_{i=1}^{n}\boldsymbol{\Sigma}-n\frac{1}{n}\boldsymbol{\Sigma}\\
&=(n-1)\boldsymbol{\Sigma}.
\end{aligned}
$$

This implies that $\mathbb{E}\,\boldsymbol{S}=\boldsymbol{\Sigma}$.

Method 2: To evaluate $\mathbb{E}(\boldsymbol{S})$, we only need to evaluate $\mathbb{E}(S_{k\ell})$ for any $1\le k,\ell,\le p$.

$$
S_{k\ell}:=\frac{1}{n-1}\sum_{i=1}^{n}(X_{ik}-\bar{X}_k)(X_{i\ell}-\bar{X}_\ell)=\frac{1}{n-1}\left(\left(\sum_{i=1}^{n}X_{ik}X_{i\ell}\right)-n\bar{X}_k\bar{X}_\ell\right).
$$

Then we have

$$
\mathbb{E}[S_{k\ell}]=\frac{1}{n-1}\left(\left(\sum_{i=1}^{n}\mathbb{E}[X_{ik}X_{i\ell}]\right)-n\,\mathbb{E}[\bar{X}_k\bar{X}_\ell]\right)
$$

Recall the formula

$$
\mathrm{Cov}(X,Y)=\mathbb{E}[XY]-\mathbb{E}[X]\,\mathbb{E}[Y]\Rightarrow\mathbb{E}[XY]=\mathrm{Cov}(X,Y)+\mathbb{E}[X]\,\mathbb{E}[Y].
$$

By

$$
\mathbb{E}[\vec{X}_i]=\mu,\quad\mathrm{Cov}(\vec{X}_i)=\boldsymbol{\Sigma}
$$

we have

$$
\mathbb{E}[X_{ik}X_{i\ell}]=\mathrm{Cov}(X_{ik},X_{i\ell})+\mathbb{E}[X_{ik}]\,\mathbb{E}[X_{i\ell}]=\sigma_{k\ell}+\mu_k\mu_\ell
$$

Similarly, by

$$
\mathbb{E}[\overline{\vec{X}}]=\mu,\quad\mathrm{Cov}(\overline{\vec{X}})=\frac{1}{n}\boldsymbol{\Sigma}
$$

we have

$$\mathbb{E}[\bar{X}_k \bar{X}_\ell] = \operatorname{Cov}(\bar{X}_k, \bar{X}_\ell) + \mathbb{E}[\bar{X}_k]\,\mathbb{E}[\bar{X}_\ell] = \frac{1}{n}\sigma_{k\ell} + \mu_k\mu_\ell$$

Therefore we have

$$\begin{aligned}
\mathbb{E}[S_{k\ell}] &= \frac{1}{n-1}\left(\left(\sum_{i=1}^{n}\mathbb{E}[X_{ik}X_{i\ell}]\right) - n\,\mathbb{E}[\bar{X}_k\bar{X}_\ell]\right) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n}(\sigma_{k\ell} + \mu_k\mu_\ell) - n\left(\frac{1}{n}\sigma_{k\ell} + \mu_k\mu_\ell\right)\right) \\
&= \sigma_{k\ell}.
\end{aligned}$$

This implies that

$$\mathbb{E}[\boldsymbol{S}] = \boldsymbol{\Sigma}.$$

$\square$

6. $n(\overline{\vec{X}} - \vec{\mu})^\top \boldsymbol{S}^{-1}(\overline{\vec{X}} - \vec{\mu}) \sim \frac{(n-1)p}{n-p} F_{p,n-p}$.

The above properties are consistent with the well-known properties for one-variate sample: Let

$$X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2).$$

with sample mean $\bar{X}$ and sample variance S^2. Then

1. $\bar{X} \sim \mathcal{N}\left(\mu, \frac{1}{n}\sigma^2\right)$;
2. $\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma} \sim \mathcal{N}(0,1)$ and $\frac{n(\overline{X}-\mu)^2}{\sigma^2} \sim \chi^2_1$;
3. $\mathbb{E}\, S^2 = \sigma^2$;
4. $\frac{\sqrt{n}(\overline{X}-\mu)}{S} \sim t_{n-1}$ and $\frac{n(\overline{X}-\mu)^2}{S^2} \sim F_{1,n-1}$.

6 Examples

Example 1: Let $\vec{X} \sim N_3(\vec{\mu}, \mathbf{\Sigma})$, where

$$\vec{\mu} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \quad \text{and} \quad \mathbf{\Sigma} = \begin{bmatrix} 3 & 1 & 0 \\ 1 & 3 & 1 \\ 0 & 1 & 2 \end{bmatrix}.$$

(a) Find the conditional distribution of $\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ given $X_3 = 0$.
(b) Find the conditional distribution of X_1 given $X_2 = X_3$.

Answer:
(a) The conditional distribution of $\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ is multivariate normal with

$$\mathbb{E}\left(\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \middle| X_3 = 0 \right) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} 2^{-1} (0 - (-1)) = \begin{bmatrix} 1 \\ \frac{1}{2} \end{bmatrix}$$

and

$$\mathrm{Cov}\left(\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \middle| X_3 = 0 \right) = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \end{bmatrix} 2^{-1} \begin{bmatrix} 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & \frac{5}{2} \end{bmatrix}.$$

(b) We know $\begin{bmatrix} X_1 \\ X_2 - X_3 \end{bmatrix}$ is multivariate normal with

$$\mathbb{E}\left(\begin{bmatrix} X_1 \\ X_2 - X_3 \end{bmatrix} \right) = \begin{bmatrix} \mathbb{E}\, X_1 \\ \mathbb{E}\, X_2 - \mathbb{E}\, X_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 - (-1) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

and

$$\mathrm{Cov}\left(\begin{bmatrix} X_1 \\ X_2 - X_3 \end{bmatrix} \right) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 3 & 1 & 0 \\ 1 & 3 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}.$$

Then we know $X_1 | X_2 - X_3 = 0$ is conditional normal. The conditional mean is

$$\mathbb{E}(X_1 | X_2 - X_3 = 0) = 1 + 1 \times \frac{1}{3} \times (0 - 1) = \frac{2}{3}$$

and the conditional variance is

$$\mathrm{Var}(X_1 | X_2 - X_3 = 0) = 3 - 1 \times \frac{1}{3} \times 1 = \frac{8}{3}.$$

Example 2: Suppose the population covariance matrix of $\vec{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ is

$$\mathbf{\Sigma} = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}.$$

We aim at finding $\mathrm{Var}(X_1 + X_2)$.

Method 1: Since $X_1 + X_2 = [1, 1]\vec{X}$, we know its population variance is

$$\mathrm{Var}(X_1 + X_2) = [1, 1] \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 7.$$

Method 2: We can avoid using matrix algebra. Notice that from $\mathbf{\Sigma} = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$, we have

$$\mathrm{Var}(X_1) = 3, \quad \mathrm{Var}(X_2) = 2, \quad \mathrm{Cov}(X_1, X_2) = 1.$$

Then

$$\begin{aligned} \mathrm{Var}(X_1 + X_2) &= \mathrm{Cov}(X_1 + X_2, X_1 + X_2) \\ &= \mathrm{Cov}(X_1, X_1) + \mathrm{Cov}(X_1, X_2) + \mathrm{Cov}(X_2, X_1) + \mathrm{Cov}(X_2, X_2). \\ &= \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\mathrm{Cov}(X_1, X_2) = 7 \end{aligned}$$

Example 3: Let

$$\vec{X} \sim \mathcal{N}_2\left(\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 5 \end{bmatrix} \right).$$

Find the marginal distributions of X_2 and $\begin{bmatrix} X_1 \\ X_3 \end{bmatrix}$.

Answer: $X_2 \sim \mathcal{N}(0, 4)$ and $\begin{bmatrix} X_1 \\ X_3 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 & 1 \\ 1 & 5 \end{bmatrix} \right)$.