Multivariate random sample matrices

In statistics, multivariate data with n observations of p-dimensional vectors

$$\begin{array}{cl}\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2k} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{j1} & x_{j2} & \cdots & x_{jk} & \cdots & x_{jp} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} & \cdots & x_{np} \end{bmatrix} & \begin{array}{l} \leftarrow \text{1st observation} \\ \leftarrow \text{2nd observation} \\ \vdots \\ \leftarrow j\text{th observation} \\ \vdots \\ \leftarrow n\text{th observation} \end{array}\end{array} \tag{1}$$

In order to study general properties of all possible data generated in similar situations, the data are considered as **observed values** or realized values of a **random sample** of p-dimensional random vectors of sample size n, which can be expressed as

$$\begin{array}{cl}\begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1k} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2k} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{j1} & X_{j2} & \cdots & X_{jk} & \cdots & X_{jp} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nk} & \cdots & X_{np} \end{bmatrix} & \begin{array}{l} \leftarrow \text{random vector, 1st draw of arandom sample} \\ \leftarrow \text{random vector, 2nd draw of arandom sample} \\ \vdots \\ \leftarrow \text{random vector, } j\text{th draw of a random sample} \\ \vdots \\ \leftarrow \text{random vector, } n\text{th draw of a random sample} \end{array}\end{array} \tag{2}$$

where every p-variate random vector is regarded as a copy of or a draw from a common p-random vector

$$[X_1\ X_2\ \cdots\ X_p]' = \begin{bmatrix} X_1 \\ \vdots \\ X_p \end{bmatrix} \tag{3}$$

which characterizes the probability distribution of the total **population** of concern. Here each component X_i is a univariate random variable.

The n draws of the random sample in (2) may or may not be dependent. In this course, unless otherwise mentioned, we assume that the n draws are mutually independent. Later on we will consider comparisons of multiple multivariate samples. At the moment, we are considering one multivariate sample from a single population, then the row random vectors in (2) are $i.i.d.$ (independent, identically distributed) copies of the common random p-vector (3) that represent the population distribution of concern.

In the following we discuss notations and properties of random vectors, random samples of p-components, and their linear transformations.

The same letter (e.g. X or $\boldsymbol{X}$) may be used to represent different variables , vectors or matrices. The meaning of shared notations should be clear from context.

1 Definitions and descriptive measures of random vectors

Analogous to the univariate case, multivariate data consists of n observations is commonly displayed in n rows and p columns, as shown in (1) (actually also commonly the other way around, i.e., displayed as p rows and n columns). As mentioned above, the data can be viewed as observed values of a sample of size n $i.i.d.$ random vectors.

To study characteristics of population distribution (3), let $\boldsymbol{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_p \end{bmatrix}$ denote a p-variate random vector.

Here, each X_j, representing a component of $\boldsymbol{X}$, is a univariate random variable.

The mean or expectation of a random vector is defined as the component-wise expectation.

$$\mathbb{E}(\boldsymbol{X}) = \begin{bmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_p \end{bmatrix} = \boldsymbol{\mu}.$$

A matrix $\boldsymbol{Y} = [Y_{ij}]_{n\times p}$ is a random matrix if its entries Y_{ij} are random variables.
Each row of $\boldsymbol{Y}$ can be viewed as the transpose of a (column) random vector.

The mean or expectation of a random matrix is defined by its element-wise mean:

$$\mathbb{E}(\mathbf{Y}) = [\mathbb{E}(Y_{ij})]_{n\times p}$$

Immediate properties of expectations of random vectors
(Exercises, using properties of the sum and product of matrices)

- $\mathbb{E}(\boldsymbol{W}+\boldsymbol{Y}) = \mathbb{E}(\boldsymbol{W}) + \mathbb{E}(\boldsymbol{Y})$, where both $\boldsymbol{W}, \boldsymbol{Y}$ are random matrices of dimension $n\times p$.
- $\mathbb{E}(A\boldsymbol{Y}B) = A\mathbb{E}(\boldsymbol{Y})B$, where $\boldsymbol{Y}$ is a random matrix of dimensions $n\times p$, A (of dimensions $k\times n$) and B $(p\times r)$ are constant matrices. The product $A\boldsymbol{Y}B$ is well defined for any integers k and r.

The covariance matrix of a p-variate random vector $\boldsymbol{X}$ is a $p\times p$ matrix Σ, sometimes indexed as Σ_x or Σ_X.

$$\Sigma = Cov(\boldsymbol{X}) = \mathbb{E}\left((\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})'\right) = \mathbb{E}\left(\begin{bmatrix} X_1-\mu_1 \\ X_2-\mu_2 \\ \vdots \\ X_p-\mu_p \end{bmatrix}\left[X_1-\mu_1,\ X_2-\mu_2,\ \cdots,\ X_p-\mu_p\right]\right)$$

By vector-matrix multiplication, the $(j,k)th$ entry of Σ is

$$\sigma_{jk} = \mathbb{E}[(X_j-\mu_j)(X_k-\mu_k)] = cov(X_j,X_k), \qquad j,k=1,\cdots,p.$$

On the diagonal of Σ,

$$\sigma_{kk} = \mathbb{E}[(X_k-\mu_k)^2] = Var(X_k) = \sigma_k^2, \qquad k=1,\cdots,p.$$

In tensor notation, we may write

$$\Sigma = Cov(\boldsymbol{X}) = \mathbb{E}\left[(\boldsymbol{X}-\boldsymbol{\mu})\otimes(\boldsymbol{X}-\boldsymbol{\mu})\right]$$

As in the case of sample covariance matrix, the determinant $|\Sigma_X| = |\Sigma| = det(\Sigma)$. sometimes is called the generalized variance of random vector $\boldsymbol{X}$.

A univariate random variable X is completely characterized by its probability distribution.
The cumulative probability distribution function of univariate X is commonly denoted as $F(x) = \mathbb{P}(X\le x)$.
If the distribution of X has mean μ and variance σ^2, we use the shorthand $X\sim(\mu,\sigma^2)$. (F maybe unknown otherwise.)

A p-variate random vector is characterized by the joint probability distribution of its components, with joint cumulative probability function $F_{\boldsymbol{X}}(x_1,\cdots,x_p) = \mathbb{P}(X_1\le x_1,\cdots,X_p\le x_p)$.

Often the mean and covariance structure are the primary interest rather than the underlying distribution F (which contains more information but can be complicated or hard to obtain).
For a random vector $\boldsymbol{X}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ, we may use the notation

$$\boldsymbol{X} \sim (\boldsymbol{\mu}, \Sigma)$$

regardless other properties of the distribution, such as whether it is multivariate normal or not.

A useful formula for covariance matrix of random vectors

Similar to the univariate case $\mathbb{E}(X^2) = \sigma^2 + \mu^2$ for $X \sim (\mu, \sigma^2)$, if $\boldsymbol{X}$ is a p-variate random vector with mean $\boldsymbol{\mu}$, then

$$\mathbb{E}(\boldsymbol{X}\boldsymbol{X}') = \Sigma + \boldsymbol{\mu}\boldsymbol{\mu}' \tag{4}$$

Proof. It is important to get the dimensions right. We provide two proofs in the following.

(1) First, we may prove the equation at the level of matrix elements (j,k), for $j,k = 1, \cdots, p$.
- Matrix $\boldsymbol{X}\boldsymbol{X}'$ is $p \times p$ with $(j,k)th$ entry $X_j X_k$.
- Matrix $\mathbb{E}(\boldsymbol{X}\boldsymbol{X}')$ is $p \times p$ and its (j,k)th entry $= \mathbb{E}(X_j X_k)$.
- Matrix $\boldsymbol{\mu}\boldsymbol{\mu}'$ is $p \times p$ and its $(j,k)th$ entry $= \mu_j \mu_k$.
- Matrix Σ is $p \times p$. Using the univariate formula, the (j,k)th entry of Σ can be written as

$$\sigma_{ij} = Cov(X_j, X_k) = \mathbb{E}(X_j X_k) - \mathbb{E}(X_j)\mathbb{E}(X_k) = \mathbb{E}(X_j X_k) - \mu_j \mu_k$$

Therefore,

the $(j,k)th$ entry of $\Sigma = \sigma_{ij} = \mathbb{E}(X_j X_k) - \mu_j\mu_k =$ the $(j,k)th$ entry of matrix $\left(\mathbb{E}(\mathbf{X}\mathbf{X}') - \boldsymbol{\mu}\boldsymbol{\mu}'\right)$

for all $j,k = 1, \cdots, p$, which implies $\Sigma = \mathbb{E}(\mathbf{X}\mathbf{X}') - \boldsymbol{\mu}\boldsymbol{\mu}'$, thus proved (4).

(2) Alternatively, we may give a proof at the matrix level. By matrix operation and by the properties of expectations of random matrices,

$$\begin{aligned}
\Sigma = Cov(\boldsymbol{X}) &= \mathbb{E}[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})'] \\
&= \mathbb{E}(\boldsymbol{X}\boldsymbol{X}') - \mathbb{E}(\boldsymbol{X})\boldsymbol{\mu}' - \boldsymbol{\mu}\mathbb{E}(\boldsymbol{X}') + \boldsymbol{\mu}\boldsymbol{\mu}' \\
&= \mathbb{E}(\boldsymbol{X}\boldsymbol{X}') - \boldsymbol{\mu}\boldsymbol{\mu}' - \boldsymbol{\mu}\boldsymbol{\mu}' + \boldsymbol{\mu}\boldsymbol{\mu}' \;=\; \mathbb{E}(\boldsymbol{X}\boldsymbol{X}') - \boldsymbol{\mu}\boldsymbol{\mu}'
\end{aligned}$$

which also proves (4).

□

2 Linear transformations of multivariate random vectors

Given a random vector $\boldsymbol{X} \in \mathbb{R}^p$, we often need to consider random vector in the form

$$\boldsymbol{Y} = A\boldsymbol{X} + \boldsymbol{c}$$

where A and c and non–random matrix and vector, respectively.

Let $\boldsymbol{X} = (X_1, \cdots, X_p)'$ be a p-variate random vector with mean $\boldsymbol{\mu}_x \in \mathbb{R}^p$ and covariance matrix $\Sigma_x \in \mathbb{R}^{p \times p}$.
Denote as $\boldsymbol{X} \sim (\boldsymbol{\mu}_x, \Sigma_x)$.

- A linear combination of the components of $\boldsymbol{X}$, written as $\boldsymbol{c}'\boldsymbol{X} = c_1 X_1 + \cdots + c_p X_p$, is a univariate random variable, with

$$\mathbb{E}(\boldsymbol{c}'\boldsymbol{X}) = \boldsymbol{c}'\boldsymbol{\mu}_x$$
$$V(\boldsymbol{c}'\boldsymbol{X}) = \boldsymbol{c}'Cov(\boldsymbol{X})\ \boldsymbol{c}$$

Both are scalars.

- If A of dimensions $k \times p$ is a matrix of constants, then $A\boldsymbol{X}$ is a k-variate random vector with k-dimensional mean vector

$$\mathbb{E}(A\boldsymbol{X}) = A\boldsymbol{\mu}_x,$$

and $k \times k$ covariance matrix

$$Cov(A\boldsymbol{X}) = ACov(\boldsymbol{X})A' = A\Sigma_x A'.$$

- More generally, if $\boldsymbol{Y}$ a linear transformation of $\boldsymbol{X}$,

$$\boldsymbol{Y} = A\boldsymbol{X} + \boldsymbol{c}$$

where A is a matrix of constants with dimensions $k \times p$, $\boldsymbol{c}$ is a k-dimensional constant vector, then the mean vector of $\boldsymbol{Y}$ is

$$\boldsymbol{\mu}_y = \mathbb{E}(\boldsymbol{Y}) = A\boldsymbol{\mu}_x + \boldsymbol{c}, \tag{5}$$

and the covariance matrix of $\boldsymbol{Y}$ is

$$\Sigma_y = Cov(\boldsymbol{Y}) = A\,Cov(\boldsymbol{X})A' = A\Sigma_x A'. \tag{6}$$

Proof. In the following we prove (5). and (6).
Notice that $A\boldsymbol{X} = A_{k\times p}\boldsymbol{X}_{p\times 1}$ and $\boldsymbol{c}$ are k-variate vectors, thus $\boldsymbol{Y} = A\boldsymbol{X} + \boldsymbol{c}$ is k-variate, and $\mathbb{E}(\boldsymbol{Y})$ is a k-vector, with jth component

$$\mathbb{E}(Y_j) = \mathbb{E}\left(\sum_{i=1}^{p} a_{ji}X_i + c_j\right) = \sum_{i=1}^{p} a_{ji}\mathbb{E}(X_i) + c_j = \ j\text{th component of } \ A\,\mathbb{E}(\boldsymbol{X}) + \boldsymbol{c}, \quad j = 1, \cdots, k.$$

Therefore $\mathbb{E}(\boldsymbol{Y}) = A\,\mathbb{E}(\boldsymbol{X}) + \boldsymbol{c}$, which proves (5).

Now consider the covariance matrix. The dimensions of $Cov(\boldsymbol{Y})$ is $k \times k$.

$$\begin{aligned}
Cov(\boldsymbol{Y}) &= \mathbb{E}[(\boldsymbol{Y} - \boldsymbol{\mu}_y)(\boldsymbol{Y} - \boldsymbol{\mu}_y)'] \\
&= \mathbb{E}\left\{[A\boldsymbol{X} + \boldsymbol{c} - (A\boldsymbol{\mu}_x + \boldsymbol{c})]\ [A\boldsymbol{X} + \boldsymbol{c} - (A\boldsymbol{\mu}_x + \boldsymbol{c})]'\right\} \\
&= \mathbb{E}[(A\boldsymbol{X} - A\boldsymbol{\mu}_x)(A\boldsymbol{X} - A\boldsymbol{\mu}_x)'] \\
&= \mathbb{E}[A(\boldsymbol{X} - \boldsymbol{\mu}_x)(A(\boldsymbol{X} - \boldsymbol{\mu}_x))'] \\
&= \mathbb{E}[A(\boldsymbol{X} - \boldsymbol{\mu}_x)(\boldsymbol{X} - \boldsymbol{\mu}_x)'A'] \\
&= A\ \mathbb{E}[(\boldsymbol{X} - \boldsymbol{\mu}_x)(\boldsymbol{X} - \boldsymbol{\mu}_x)']A') \\
&= ACov(\boldsymbol{X})A' = A\Sigma_x A'
\end{aligned}$$

which proves (6).
In the steps of the above proof, we applied matrix operation properties that $(AB)' = B'A'$ for matrices A, B, and that $\mathbb{E}(A\boldsymbol{X}B) = A\,\mathbb{E}(\boldsymbol{X})B$ for scalar (non-random) matrices A, B and a random matrix $\boldsymbol{X}$.

□

3 Random sample matrix

We now consider the point of view that data are samples generated from multivariate random variables.

Denote the jth observation of the kth variable as x_{jk}, $k = 1, \dots, p$, $j = 1 \cdots, n$.

Before the measurement is taken, there are many possible values for x_{jk}. The before-measurement sample value of the jth observation of the kth variable is a random variable, denoted as X_{jk}. The before-measurement possible outcome of the jth observation on all p variables is a p-variate **random vector**, denoted as $\boldsymbol{X}_j = (X_{j1}, \cdots, X_{jp})'$.

When we consider the general properties of a sample of n observations before the measurements are taken, we are examining a collection of n random vectors following the same joint p-variate distribution.

Let random p-vectors $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ be a random sample with mean $\boldsymbol{\mu}$ and covariance matrix Σ_x.

Unless otherwise stated, by random sample we mean that $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ are n **independent** p-variate observations, and they from a **common** p-variate distribution of mean $\boldsymbol{\mu}$ and covariance Σ_x in the one sample case we considered in this section.

$$\boldsymbol{X} = [X_{jk}]_{np} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1k} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2k} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{j1} & X_{j2} & \cdots & X_{jk} & \cdots & X_{jp} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nk} & \cdots & X_{np} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1' \\ \boldsymbol{X}_2' \\ \vdots \\ \boldsymbol{X}_j' \\ \vdots \\ \boldsymbol{X}_n' \end{bmatrix} \tag{7}$$

Remarks on random sample matrix (7)

- On the independence assumption
  There are many data applications that the sample observations (row vectors) are not independent, e.g., time series or longitudinal observations, or repeated observations from the same subjects.
  Without specific statement, in this course, observations in a ransom sample, that is, the row vectors in (7), are assumed independent.
  Further more, every row vector is of the same distribution when we talk about one sample.
- Random sample matrix vs general random matrix
  Although (7) is a random matrix, we use it chiefly for the simple random sample case, so the n random sample vectors $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ are independent vectors of the same p-variate joint distribution. Theerfore random sample matrix is a simple, special case of random matrices. Beyond the element-wise means, Its other descriptive properties, such as $Cov(\boldsymbol{X}) = \boldsymbol{I}_n \otimes \Sigma$, would call for tensor methods and other concepts that won't be discussed in this course. The study of general random matrices is a vast field.
- Analogous to univariate analysis, the quantities of interest are the sample average which is p-component vector, and sample covariance of the p-component variables which is a $p \times p$ matrix.

For sample random vectors $\boldsymbol{X}_i, i = 1, \cdots, n$ in (7), the **sample mean** is a random p-vector

$$\bar{\boldsymbol{X}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i$$

and the **sample covariance** (a $p \times p$ random matrix) is defined as

$$\boldsymbol{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})'$$

or

$$\boldsymbol{S}_n = \frac{n-1}{n}\boldsymbol{S} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})'$$

The two definitions of covariance ($\boldsymbol{S}$ and $\boldsymbol{S}_n$) are for usage convenience. While $\boldsymbol{S}$ is an unbiased estimator of the population covariance matrix Σ (as shown below), $\boldsymbol{S}_n$ can be convenient in derivations.

Under the assumption that the sample vectors are i.i.d. copies of the same distribution,

$$\boldsymbol{X}_i \sim (\boldsymbol{\mu}, \Sigma)$$

we can show in the following that, that mean and covariance for the mean vector $\bar{\boldsymbol{X}}$ are

$$\mathbb{E}(\bar{\boldsymbol{X}}) = \boldsymbol{\mu}, \qquad Cov(\bar{\boldsymbol{X}}) = \frac{1}{n}\Sigma_x, \qquad \mathbb{E}(\boldsymbol{S}_n) = \frac{n}{n-1}\Sigma_x. \tag{8}$$

Proof. In the following we derive the three expressions in (8).

(i) Derivation of $E(\bar{\boldsymbol{X}}) = \boldsymbol{\mu}$

$$\mathbb{E}(\bar{\boldsymbol{X}}) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(\boldsymbol{X}_i) = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\mu} = \boldsymbol{\mu}$$

(ii) Derivation of $Cov(\bar{\boldsymbol{X}}) = \frac{1}{n}\Sigma_x$

Step 1 – Expand $Cov(\bar{\boldsymbol{X}}) = \mathbb{E}\left[(\bar{\boldsymbol{X}} - \boldsymbol{\mu})(\bar{\boldsymbol{X}} - \boldsymbol{\mu})'\right]$ as a sum of n^2 matrices.

The covariance of random vector $\bar{\boldsymbol{X}}$ is the $p \times p$ matrix

$$Cov(\bar{\boldsymbol{X}}) = \mathbb{E}\left[(\bar{\boldsymbol{X}} - \boldsymbol{\mu})(\bar{\boldsymbol{X}} - \boldsymbol{\mu})'\right].$$

To show that the covariance matrix has the form in (8), we rewrite

$$\bar{\boldsymbol{X}} - \boldsymbol{\mu} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i - \boldsymbol{\mu} = \frac{1}{n}\left(\sum_{i=1}^{n}\boldsymbol{X}_i - n\boldsymbol{\mu}\right) = \frac{1}{n}\left(\sum_{i=1}^{n}\boldsymbol{X}_i - \sum_{i=1}^{n}\boldsymbol{\mu}\right) = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X}_i - \boldsymbol{\mu}).$$

The $p \times p$ matrix

$$(\bar{\boldsymbol{X}} - \boldsymbol{\mu})(\bar{\boldsymbol{X}} - \boldsymbol{\mu})' = \left(\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X}_i - \boldsymbol{\mu})\right)\left(\frac{1}{n}\sum_{j=1}^{n}(\boldsymbol{X}_j - \boldsymbol{\mu})\right)' = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\boldsymbol{X}_i - \boldsymbol{\mu})(\boldsymbol{X}_j - \boldsymbol{\mu})',$$

which is the summation of n^2 random matrices, each of the form

$$(\boldsymbol{X}_i - \boldsymbol{\mu})(\boldsymbol{X}_j - \boldsymbol{\mu})' = \begin{bmatrix} X_{i1} - \mu_1 \\ \vdots \\ X_{ip} - \mu_p \end{bmatrix}\left[X_{j1} - \mu_1, \ \cdots, \ X_{jp} - \mu_p\right].$$

Step 2 – Show that only n of the n^2 matrix means are not the zero matrix.

When $i \neq j$, any pairs of components X_{ik} of $\boldsymbol{X}_i$ and $X_{j\ell}$ of $\boldsymbol{X}_j$ are independent, because they are components of two independent observations $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$.
Consequently, the univariate component variables X_{ik} and $X_{j\ell}$ are uncorrelated, which means

$$cov(X_{ik}, X_{j\ell}) = \mathbb{E}(X_{ik} - \mu_k)(X_{j\ell} - \mu_\ell) = 0, \qquad for \quad i \neq j \quad and\ any \quad k, \ell = 1, \cdots, n.$$

Within the ith observation $\boldsymbol{X}_i$, the covariance of the kth and ℓth components X_{ik} and $X_{i\ell}$ is the covariance of the original kth and ℓth variables of the common joint distribution that generated the random sample. That is,

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

which is the (k, ℓ)th entry of Σ_x.
Therefore, the expectation of the $p \times p$ random matrix $(\boldsymbol{X}_i - \boldsymbol{\mu})(\boldsymbol{X}_j - \boldsymbol{\mu})'$ has the form

$$\mathbb{E}\left[(\boldsymbol{X}_i - \boldsymbol{\mu})(\boldsymbol{X}_j - \boldsymbol{\mu})'\right] = \begin{cases} [\mathbb{E}[(X_{ik} - \mu_k)(X_{j\ell} - \mu_\ell)]]_{k,\ell=1,\cdots,p} = [0]_{k,\ell=1,\cdots,p} & = \boldsymbol{O}_{p\times p}, \qquad \forall i \neq j. \\ [\mathbb{E}[(X_{ik} - \mu_k)(X_{i\ell} - \mu_\ell)]]_{k,\ell=1,\cdots,p} = [\sigma_{k\ell}]_{k,\ell=1,\cdots,p} & = \Sigma_x, \qquad \forall i = j. \end{cases}$$

Thus the double sum

$$\sum_{i=1}^n \sum_{j=1}^n \mathbb{E}\left[(\boldsymbol{X}_i - \boldsymbol{\mu})(\boldsymbol{X}_j - \boldsymbol{\mu})'\right] = \sum_{i=1}^n \mathbb{E}\left[(\boldsymbol{X}_i - \boldsymbol{\mu})(\boldsymbol{X}_i - \boldsymbol{\mu})'\right] = n\,\Sigma_x$$

<u>Step 3</u> – Arrive at $Cov(\bar{\boldsymbol{X}}) = \Sigma_x/n$.

Consequently, the covariance of $\bar{\boldsymbol{X}}$ can be written as

$$\begin{aligned} Cov(\bar{\boldsymbol{X}}) =& \mathbb{E}\left[(\bar{\boldsymbol{X}} - \boldsymbol{\mu})(\bar{\boldsymbol{X}} - \boldsymbol{\mu})'\right] \\ =& \mathbb{E}\left(\frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n (\boldsymbol{X}_i - \boldsymbol{\mu})(\boldsymbol{X}_j - \boldsymbol{\mu})'\right) \\ =& \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n \mathbb{E}[(\boldsymbol{X}_i - \boldsymbol{\mu})(\boldsymbol{X}_j - \boldsymbol{\mu})'] \\ =& \frac{1}{n^2}\sum_{i=1}^n \mathbb{E}\left[(\boldsymbol{X}_i - \boldsymbol{\mu})(\boldsymbol{X}_i - \boldsymbol{\mu})'\right] = \frac{1}{n^2} n\,\Sigma_x = \frac{1}{n}\Sigma_x \end{aligned}$$

where the last three equalities are from the last equation of Step 2.

(iii) Derivation of $E(\boldsymbol{S}_n) = \frac{n}{n-1}\Sigma_x$

Finally we get to validate the expression for the mean of the sample covariance matrix $\boldsymbol{S}_n$.

$$\begin{aligned} \mathbb{E}(\boldsymbol{S}_n) &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}[(\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})'] \\ &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left(\boldsymbol{X}_i\boldsymbol{X}_i' - \boldsymbol{X}_i\bar{\boldsymbol{X}}' - \bar{\boldsymbol{X}}\boldsymbol{X}_i{}' + \bar{\boldsymbol{X}}\bar{\boldsymbol{X}}'\right) \\ &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}(\boldsymbol{X}_i\boldsymbol{X}_i') - \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i\right)\bar{\boldsymbol{X}}'\right] - \mathbb{E}\left[\bar{\boldsymbol{X}}\left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i{}'\right)\right] + \frac{1}{n}\sum_{i=1}^n \mathbb{E}(\bar{\boldsymbol{X}}\bar{\boldsymbol{X}}') \end{aligned}$$

Applying $(A+B)' = A' + B'$ to the last expression,

$$\begin{aligned} \mathbb{E}(\boldsymbol{S}_n) &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}(\boldsymbol{X}_i\boldsymbol{X}_i') - \mathbb{E}(\bar{\boldsymbol{X}}\bar{\boldsymbol{X}}') - E(\bar{\boldsymbol{X}}\bar{\boldsymbol{X}}') + \frac{1}{n}n\,\mathbb{E}(\bar{\boldsymbol{X}}\bar{\boldsymbol{X}}') \\ &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}(\boldsymbol{X}_i\boldsymbol{X}_i') - \mathbb{E}(\bar{\boldsymbol{X}}\bar{\boldsymbol{X}}') \end{aligned}$$

Use the notations $\mathbb{E}(\boldsymbol{X}_i) = \boldsymbol{\mu}, Cov(\boldsymbol{X}_i) = \Sigma_x$ for $i = 1, \cdots, n$,

$$\mathbb{E}(\boldsymbol{S}_n) \;=\; \frac{1}{n}\sum_{i=1}^n (\Sigma_x + \boldsymbol{\mu}\boldsymbol{\mu}') - \left(\frac{1}{n}\Sigma_x + \boldsymbol{\mu}\boldsymbol{\mu}'\right) \;=\; \frac{n-1}{n}\Sigma_x$$

This concludes the derivations of the expressions in (8):

$$\mathbb{E}(\bar{\boldsymbol{X}}) = \boldsymbol{\mu}, \qquad Cov(\bar{\boldsymbol{X}}) = \frac{1}{n}\Sigma_x, \qquad \mathbb{E}(\boldsymbol{S}_n) = \frac{n}{n-1}\Sigma_x.$$

□

<u>Note</u> Relevant chapters in the text by Johnson and Wichern: Chapter 2 and 3.