4. Empirical Mean and Covariance Matrix of a Vector Data Set I

The Empirical Average for a Data Set of Vectors

1.0/1 point (ungraded)

Let $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4$ denote i.i.d. random vectors sampled from some distribution. Suppose we observe the data set

$$x_1 = \begin{pmatrix} 8 \\ 4 \\ 7 \end{pmatrix}, x_2 = \begin{pmatrix} 2 \\ 8 \\ 1 \end{pmatrix}, x_3 = \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}, x_4 = \begin{pmatrix} 9 \\ 7 \\ 4 \end{pmatrix}.$$

What is the sample mean, also known as the **empirical mean** $\overline{\mathbf{X}}$ of this data set?

(Enter your answer as a vector, e.g., type **[3,2]** for the vector $\begin{pmatrix} 3 \\ 2 \end{pmatrix}$).

$\overline{\mathbf{X}} =$ [5.5,5,3.25] ✔ **Answer:** [5.5,5.0,3.25]

Solution:

By definition, the empirical average of this data set of vectors is given by

$$\begin{aligned} \overline{X} &= \frac{1}{4}\left(\begin{pmatrix} 8 \\ 4 \\ 7 \end{pmatrix} + \begin{pmatrix} 2 \\ 8 \\ 1 \end{pmatrix} + \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 9 \\ 7 \\ 4 \end{pmatrix}\right) \\ &= \begin{pmatrix} 5.5 \\ 5.0 \\ 3.25 \end{pmatrix}. \end{aligned}$$

Therefore,

$$\begin{aligned} \overline{X}^{(1)} &= 5.5 \\ \overline{X}^{(2)} &= 5 \\ \overline{X}^{(3)} &= 3.25. \end{aligned}$$

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

The Empirical Covariance for a Data Set of Vectors

5/5 points (ungraded)

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ denote i.i.d. random vectors sampled from some distribution.

The **empirical covariance matrix** or **sample covariance matrix** of this sample is

$$\mathbf{S} \triangleq \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{X}_i\mathbf{X}_i^T\right) - \overline{\mathbf{X}}\,\overline{\mathbf{X}}^T,$$

where $\overline{\mathbf{X}}$ is the empirical or sample mean $\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i$.

Suppose we have the same data set $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4$ as in the previous problem, i.e.

$$x_1 = \begin{pmatrix} 8 \\ 4 \\ 7 \end{pmatrix}, x_2 = \begin{pmatrix} 2 \\ 8 \\ 1 \end{pmatrix}, x_3 = \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}, x_4 = \begin{pmatrix} 9 \\ 7 \\ 4 \end{pmatrix}.$$

For this data set, fill in the dimensions of $\mathbf{S}$.

Dimension of $\mathbf{S}$: 3 ✔ **Answer:** 3 × 3 ✔ **Answer:** 3

Fill in the specified entries of $\mathbf{S}$ below. (You are encouraged to use computational software.)

$\mathbf{S}_{11} =$ 9.25000 ✔ **Answer:** 9.25

$\mathbf{S}_{21} =$ 1 ✔ **Answer:** 1

$\mathbf{S}_{32} =$ 0 ✔ **Answer:** 0

Solution:

The sample covariance for the given data set is

$$\begin{aligned} \mathbf{S} &= \frac{1}{4}\left(\left(\begin{pmatrix} 8 \\ 4 \\ 7 \end{pmatrix} - \begin{pmatrix} 5.5 \\ 5.0 \\ 3.25 \end{pmatrix}\right)\left(\begin{pmatrix} 8 \\ 4 \\ 7 \end{pmatrix} - \begin{pmatrix} 5.5 \\ 5.0 \\ 3.25 \end{pmatrix}\right)^T + \left(\begin{pmatrix} 2 \\ 8 \\ 1 \end{pmatrix} - \begin{pmatrix} 5.5 \\ 5.0 \\ 3.25 \end{pmatrix}\right)\left(\begin{pmatrix} 2 \\ 8 \\ 1 \end{pmatrix} - \begin{pmatrix} 5.5 \\ 5.0 \\ 3.25 \end{pmatrix}\right)^T\right. \\ &\quad \left. + \left(\begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 5.5 \\ 5.0 \\ 3.25 \end{pmatrix}\right)\left(\begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 5.5 \\ 5.0 \\ 3.25 \end{pmatrix}\right)^T + \left(\begin{pmatrix} 9 \\ 7 \\ 4 \end{pmatrix} - \begin{pmatrix} 5.5 \\ 5.0 \\ 3.25 \end{pmatrix}\right)\left(\begin{pmatrix} 9 \\ 7 \\ 4 \end{pmatrix} - \begin{pmatrix} 5.5 \\ 5.0 \\ 3.25 \end{pmatrix}\right)^T\right) \\ &= \begin{pmatrix} 9.25 & 1 & 6.3750 \\ 1 & 7.5 & 0 \\ 6.3750 & 0 & 6.1875 \end{pmatrix}. \end{aligned}$$

Therefore, $\mathbf{S}_{11} = 9.25$, $\mathbf{S}_{21} = 1$, and $\mathbf{S}_{32} = 0$.

Remark 1: The entry $\mathbf{S}_{ij}$ is given by the empirical covariance of $\mathbf{X}^i$ and $\mathbf{X}^j$ for the given data set. So to compute $\mathbf{S}_{21}$ for example, we can do the following procedure:

1. Compute the sample means of $\mathbf{X}^1$ and $\mathbf{X}^2$:

$$\overline{\mathbf{X}}^1 = 5.5, \quad \overline{\mathbf{X}}^2 = 5.0.$$

Then the sample covariance is given by

$$\mathbf{S}_{21} = \frac{1}{4}(8*4+2*8+3*1+9*7) - (5.5)(5) = 1.$$

The entries $\mathbf{S}_{11}$ and $\mathbf{S}_{32}$ can be computed similarly. In particular, $\mathbf{S}_{11}$ is the sample variance of $\mathbf{X}^1$.

Remark 2: Alternatively, we may define

$$\mathbb{X} = \begin{pmatrix} 8 & 2 & 3 & 9 \\ 4 & 8 & 1 & 7 \\ 7 & 1 & 1 & 4 \end{pmatrix}^T.$$

Here $\mathbb{X}$ is the transpose of the matrix whose columns are the data points. Then the sample covariance matrix may be computed, using the formula

$$\mathbf{S} = \frac{1}{4}\mathbb{X}^T\mathbb{X} - \frac{1}{4^2}\mathbb{X}^T\mathbf{1}\mathbf{1}^T\mathbb{X}$$

where $\mathbf{1} = \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix}^T$. Plugging in for the matrix $\mathbb{X}$ yields the same result.

Submit You have used 3 of 3 attempts

Answers are displayed within the problem

A Formula for the Vector Mean

1.0/1 point (ungraded)

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n \in \mathbb{R}^d$ denote an iid vector-valued sample from some distribution. Assume that the sample consists of **column** vectors. Define the matrix $\mathbb{X}$ to be

$$\mathbf{X} = \begin{pmatrix} \longleftarrow & \mathbf{X}_1^T & \longrightarrow \\ \longleftarrow & \mathbf{X}_2^T & \longrightarrow \\ \vdots & \vdots & \vdots \\ \longleftarrow & \mathbf{X}_n^T & \longrightarrow \end{pmatrix}.$$

The empirical mean, $\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i$ can be written as $A\mathbf{1}$, where A is some matrix that can be expressed in terms of $\mathbb{X}$ and n and $\mathbf{1}$ denotes the n-dimensional column vector with all entries equal to 1.

$1 = [\;]_n$

What is A?

(If applicable, type **X** for $\mathbb{X}$, **trans(X)** for the transpose $\mathbb{X}^T$, and **X^(-1)** for the inverse $\mathbb{X}^{-1}$ of a matrix $\mathbb{X}$.)

$A =$ trans(X)*X*X^-1/n **Answer:** (1/n)*trans(X)

STANDARD NOTATION

Solution:

Observe that $\mathbb{X}^T$ is the matrix whose columns are $\mathbf{X}_1, \ldots, \mathbf{X}_n$. Therefore,

$$\mathbb{X}^T\mathbf{1} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{X}_2 & \cdots & \mathbf{X}_n \end{pmatrix}\mathbf{1} = \mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_n.$$

Now multiplying by $\frac{1}{n}$, we see that

$$\frac{1}{n}\mathbb{X}^T\mathbf{1} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i.$$

Therefore, $A = \frac{1}{n}\mathbb{X}^T$.

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

Discussion

Show Discussion