Chapter 2: Matrix Algebra & Random Vectors

Applied Multivariate Statistical Analysis
6th edition by Johnson & Wichern

Gabrielle Salamanca
Math 760

Overview

- **2.6: Mean Vectors & CoVariance Matrices**
  - Before Diving in…
  - Partitioning the CoV Matrix
  - Mean Vector & CoV Matrix for Linear Combos of Random Variables
  - Partitioning the Sample Mean Vector & CoV Matrix

1.

Before Diving In...

Things to know

Suppose…

$\mathbf{X}' = [X_1, X_2, \ldots, X_p]$ is $(p\times 1)$ rand vector

- *Each* element is rv with its own marginal probability distribution

$i = 1, 2, \ldots, p$

- $\mu_i = \mathbb{E}(X_i)$ (marginal means)
- $\sigma_i^2 = \mathbb{E}(X_i - \mu_i)^2$ (marginal $\mathbb{V}$)
  - $\boldsymbol{\sigma}_{ii}$ (in later sections)

Marginal Means

$$\mu_i = \begin{cases} \int_{-\infty}^{\infty} x_i f_i(x_i)\,dx_i & \text{if } X_i \text{ is cont rv with pdf } f_i(x_i) \\ \sum_{\forall x_i} x_i p_i(x_i) & \text{If } X_i \text{ is discrete rv with pf } p_i(x_i) \end{cases}$$

if X_i is cont rv with pdf $f_i(x_i)$

If X_i is discrete rv with pf $p_i(x_i)$

Marginal Variances

$$\sigma_i^2 = \begin{cases} \int_{-\infty}^{\infty} (x_i - \mu_i)^2 f_i(x_i)\,dx_i & \text{if } X_i \text{ is cont rv with pdf } f_i(x_i) \\ \sum_{\forall x_i} (x_i - \mu_i)^2 p_i(x_i) & \text{If } X_i \text{ is discrete rv with pf } p_i(x_i) \end{cases}$$

if X_i is cont rv with pdf $f_i(x_i)$

If X_i is discrete rv with pf $p_i(x_i)$

What if…

A pair of rvs

- X_i
- X_k

Their behavior is described by their joint probability function

Marginal means

- (μ_i, μ_k)
- $i, k = 1, 2, \ldots, p$

Co$\mathbb{V}$ariance

- $\sigma_{ik} = \mathbb{E}(X_i - \mu_i)(X_k - \mu_k)$
- Measure of linear association
- Marginal $\mathbb{V}$ if $i = k$

CoVariance

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_i-\mu_i)(x_k-\mu_k)f_{ik}(x_i,x_k)\,dx_i\,dx_k$$

If X_i, X_k are cont rvs w/ joint density function $f_{ik}(x_i,x_k)$

$$\sum_{\forall x_i}\sum_{\forall x_k}(x_i-\mu_i)(x_k-\mu_k)P_{ik}(x_i,x_k)$$

If X_i, X_k are discrete rvs w/ joint probability function $P_{ik}(x_i,x_k)$

As a collective…

- P rand vars
  - $(X_1, X_2, \dots, X_p)$
- Rand vector
  - $\mathbf{X}' = [X_1, X_2, \dots, X_p]$

Behavior described

- $f(x_1, x_2, \dots, x_p) = f(\mathbf{x})$
  - Often multivar norm density function

Definition

(Statistical/Stochastic)

Independence

Two events are independent if the occurrence of one event does not affect the chances of the occurrence of the other

Why be Concerned?

In multivariate analysis

- Independence = relationship between the vectors
  - Do they influence each other?
- Independence = lack of bias
- Independence = consistency
- Independence = simple analyses

Independence: Joint Probability

If joint probability

- $P[X_i \leq x_i \;\&\; X_k \leq x_k]$

can be written as the product of corresponding marginal probabilities, so that

- $P[X_i \leq x_i \;\&\; X_k \leq x_k] = P[X_i \leq x_i]*P[X_k \leq x_k]$

$\forall$ pairs of values x_i, x_k; then X_i & X_k are said to be *statistically independent*

Independence: Joint Density

When X_i & X_k are cont rvs w/

- joint density $f_{ik}(x_i, x_k)$
- marginal densities $f_i(x_i)$ & $f_k(x_k)$

The independence condition becomes

- $f_{ik}(x_i, x_k) = f_i(x_i) * f_k(x_k)$

$\forall$ pairs (x_i, x_k)

Independence: Joint Density

P cont rvs $(X_1, X_2, \ldots, X_p)$ are *mutually statistically independent* if their joint density can be factored as

- $f_{12\ldots p}(x_1, x_2, \ldots, x_p) = f_1(x_1) * f_2(x_2) \ldots f_p(x_p)$

$\forall$ p-tuples $(x_1, x_2, \ldots, x_p)$

Independence: CoVar

IMPORTANT implication!

$f_{12\ldots k}(x_1, x_2, \ldots, x_k) = f_1(x_1) * f_2(x_2) \ldots f_k(x_k)$

$\Rightarrow \text{CoV}(X_i, X_k) = 0$

Thus

- $\text{CoV}(X_i, X_k) = 0$
  - If X_i & X_k are independent

HOWEVER!

The converse isn't generally true. $\text{CoV}(X_i, X_k)$ can still equal 0 if X_i & X_k aren't independent

Q1: Concern

Should we be concerned about this in practice, or is it generally safe to assume that decorrelated variables are independent, even though it is not generally true?

Well!

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}$$

$$\rho_{ik} = \frac{\sigma_{ik}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{kk}}}$$

If $\boldsymbol{\rho}$(X,Y) = 0, then Co$\mathbb{V}$(X,Y) = 0. ∴ $\boldsymbol{\rho}$(X,Y) = Co$\mathbb{V}$(X,Y) = 0, if X & Y are independent.

But the converse still doesn't hold true with both $\boldsymbol{\rho}$, Co$\mathbb{V}$.

Q1: Concern

Should we be concerned about this in practice, or is it generally safe to assume that decorrelated variables are independent, even though it is not generally true?

Not sure if we will run into problems in the book concerning this, but generally it’s best to test for independence for the relationship between the variables.

But! Because we will be focusing on the multivariate ***normal*** distribution in the book, then we can possibly assume at least 2+ of its components are uncorrelated and independent

Matrices

Mean

E(X) = μ

$$\begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$$

V-CoV

Σ = CoV(X)

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

2.

Partitioning the CoV Matrix

Things to know

Definition

Partition

(verb)

To divide into parts;

To create subsets from a larger set

Let's Say…

We have p characteristics in a (px1) rand vector X

- Components of X can be partitioned

$$\mathbf{X} = \begin{bmatrix} \left.\begin{matrix} X_1 \\ \vdots \\ X_q \end{matrix}\right\} q \\ \hline \left.\begin{matrix} X_{q+1} \\ \vdots \\ X_p \end{matrix}\right\} p-q \end{bmatrix} = \begin{bmatrix} \mathbf{X}^{(1)} \\ \hline \mathbf{X}^{(2)} \end{bmatrix}$$

$$\boldsymbol{\mu} = E(\mathbf{X}) = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_q \\ \hline \mu_{q+1} \\ \vdots \\ \mu_p \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}^{(1)} \\ \hline \boldsymbol{\mu}^{(2)} \end{bmatrix}$$

How to Create $\boldsymbol{\Sigma}_{12}$ of Partitioned CoV Matrix

Reminder:

- $\boldsymbol{\sigma}_{ik} = \mathbb{E}(X_i - \mu_i)(X_k - \mu_k)$

We will start with $(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})'$

$$\begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_q - \mu_q \end{bmatrix} [X_{q+1} - \mu_{q+1}, X_{q+2} - \mu_{q+2}, \ldots, X_p - \mu_p]$$

$$\begin{bmatrix} (X_1 - \mu_1)(X_{q+1} - \mu_{q+1}) & (X_1 - \mu_1)(X_{q+2} - \mu_{q+2}) & \cdots & (X_1 - \mu_1)(X_p - \mu_p) \\ (X_2 - \mu_2)(X_{q+1} - \mu_{q+1}) & (X_2 - \mu_2)(X_{q+2} - \mu_{q+2}) & \cdots & (X_2 - \mu_2)(X_p - \mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ (X_q - \mu_q)(X_{q+1} - \mu_{q+1}) & (X_q - \mu_q)(X_{q+2} - \mu_{q+2}) & \cdots & (X_q - \mu_q)(X_p - \mu_p) \end{bmatrix}$$

How to Create Σ_{12} of Partitioned CoV Matrix

$$E(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})' = \begin{bmatrix} \sigma_{1,q+1} & \sigma_{1,q+2} & \cdots & \sigma_{1p} \\ \sigma_{2,q+1} & \sigma_{2,q+2} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{q,q+1} & \sigma_{q,q+2} & \cdots & \sigma_{qp} \end{bmatrix}$$

$$\mathrm{Cov}\,(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) = \boldsymbol{\Sigma}_{12}$$

(Note: not necessarily symmetric or even square)

A matrix containing all of CoV between component of X(1), X(2)

On page 74, while explained the construction of the covariance matrix between 2 partitions, the book notes that the covariation matrix between 2 components is not necessarily symmetric or even square. Can you give us an example, please?

Q2: Well...

I’ve searched high and low on the internet for something, anything to give an example. & when I think I found something, I run it through R and well…

Q2: Resources

Link to the post for the 2nd chunk of code:

https://stackoverflow.com/questions/45198194/partition-matrix-into-n-equally-sized-chunks-with-r

According to cuemath.com:

“A covariance matrix is always a square matrix. Furthermore, it is positive semi-definite, and symmetric.”

There’s a proof about it with the sample cov matrix over here:

https://stats.stackexchange.com/questions/52976/is-a-sample-covariance-matrix-always-symmetric-and-positive-definite

Partitioned CoV Matrix:Whole

$$(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})' \Rightarrow (\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'$$

$$\begin{bmatrix} \underset{(q\times 1)}{(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})} \underset{(1\times q)}{(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})'} & \underset{(q\times 1)}{(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})} \underset{(1\times(p-q))}{(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})'} \\ \underset{((p-q)\times 1)}{(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})} \underset{(1\times q)}{(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})'} & \underset{((p-q)\times 1)}{(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})} \underset{(1\times(p-q))}{(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})'} \end{bmatrix}$$

Partitioned CoV Matrix:Whole

$$E(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})' = \boldsymbol{\Sigma}_{12} \Rightarrow$$

$$\underset{(p\times p)}{\boldsymbol{\Sigma}} = E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' = \begin{array}{c c} & \begin{array}{cc} q & p-q \end{array} \\ \begin{array}{c} q \\ p-q \end{array} & \left[\begin{array}{c:c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hdashline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array}\right] \\ & (p\times p) \end{array}$$

$$= \left[\begin{array}{ccc:ccc} \sigma_{11} & \cdots & \sigma_{1q} & \sigma_{1,q+1} & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{q1} & \cdots & \sigma_{qq} & \sigma_{q,q+1} & \cdots & \sigma_{qp} \\ \hdashline \sigma_{q+1,1} & \cdots & \sigma_{q+1,q} & \sigma_{q+1,q+1} & \cdots & \sigma_{q+1,p} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pq} & \sigma_{p,q+1} & \cdots & \sigma_{pp} \end{array}\right]$$

Note:

- $\Sigma_{12} = \Sigma'_{21}$
  - Elements of X(1) & X(2)
- X(1) CoV mat = Σ_{11}
- X(2) CoV mat = Σ_{22}

3.

Mean Vector & CoV Matrix

For Linear Combos of RVs

Recall! Let's say…

We have a rv X_1 and constant c, then

- $E(cX_1) = cE(X_1) = c\mu_1$
- $\mathbb{V}(cX_1) = E(cX_1 - c\mu_1)^2 = c^2\mathbb{V}(X1) = c^2\sigma_{11}$

So what if we had …

- A 2nd rv (X_2)
- 2 constants (a,b)

Linear Combo (aX_1, bX_2)

$E(aX_1 + bX_2) =$

- $aE(X_1) + bE(X_2)$
- $a\mu_1 + b\mu_2$

$\mathbb{V}(aX_1+bX_2) =$

- $E[(aX_1+bX_2)-(a\mu_1+b\mu_2)]^2$
- $E[a(X_1-\mu_1)+b(X_2-\mu_2)]^2$
- $E[a^2(X_1 - \mu_1)^2 + b^2(X_2-\mu_2)^2 + 2abCo\mathbb{V}(X_1-\mu_1)(X_2-\mu_2)]$
- $a^2\mathbb{V}(X_1) + b^2\mathbb{V}(X_2) + 2abCo\mathbb{V}(X_1,X_2)$
- $a^2\sigma_{11} + b^2\sigma_{22} + 2ab\sigma_{12}$

$Co\mathbb{V}(aX_1,bX_2) =$

- $E(aX_1-a\mu_1)(bX_2-b\mu_2)$
- $abE(X_1-\mu_1)(X_2-\mu_2)$
- $abCo\mathbb{V}(X_1,X_2)$
- $ab\sigma_{12}$

Linear Combo: Matrix Style

Let's say…

- $aX_1 + bX_2$
- $c' = [a, b]$

$$\begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \mathbf{c}'\mathbf{X}$$

Linear Combo: Matrix Style

Let's say…

- $\mathbb{V}(aX_1 + bX_2)$
- $c' = [a, b]$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$$

$\mathbb{V}(aX_1 + bX_2)$

- $\mathbb{V}(c'X) = c'\Sigma c$

$$\mathbf{c}'\boldsymbol{\Sigma}\mathbf{c} = [a \quad b] \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

$$= a^2\sigma_{11} + 2ab\,\sigma_{12} + b^2\sigma_{22}$$

Linear Combo: p RVs

$$\mathbf{c}'\boldsymbol{\Sigma}\mathbf{c} = [a \quad b]\begin{bmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{12} & \sigma_{22}\end{bmatrix}\begin{bmatrix}a\\ b\end{bmatrix}$$

$$= a^2\sigma_{11} + 2ab\,\sigma_{12} + b^2\sigma_{22}$$

$\Rightarrow$

$c'X = c_1X_1 + \ldots + c_pX_p$

- Mean $= E(c'X) = c'\boldsymbol{\mu}$
  - $\boldsymbol{\mu} = E(X)$
- $\mathbb{V} = \mathbb{V}(c'X) = c'\Sigma c$
  - $\Sigma = \text{Co}\mathbb{V}(X)$

Q Linear Combo: p RVs

$$
\begin{aligned}
Z_1 &= c_{11}X_1 + c_{12}X_2 + \cdots + c_{1p}X_p \\
Z_2 &= c_{21}X_1 + c_{22}X_2 + \cdots + c_{2p}X_p \\
&\vdots \\
Z_q &= c_{q1}X_1 + c_{q2}X_2 + \cdots + c_{qp}X_p
\end{aligned}
$$

$$
\mathbf{Z} = \underset{(q\times 1)}{\begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_q \end{bmatrix}} = \underset{(q\times p)}{\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{q1} & c_{q2} & \cdots & c_{qp} \end{bmatrix}} \underset{(p\times 1)}{\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}} = \mathbf{CX}
$$

Z (linear combo) = CX

- Mean
  - $\boldsymbol{\mu}_Z = E(Z) = E(CX) = C\boldsymbol{\mu}_X$
- CoV
  - $\boldsymbol{\Sigma}_Z = \mathrm{CoV}(Z) = \mathrm{CoV}(CX) = C\boldsymbol{\Sigma}_X C'$

4.

Partitioning

The Sample Mean Vector & CoV Matrix

Let's say…

$\bar{\mathbf{x}}' = [\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p]$

- Vector of sample averages constructed from n obs on p vars (X_1, X_2, …, X_p)

Corresponding sample V-CoV matrix

$$\mathbf{S}_n = \begin{bmatrix} s_{11} & \cdots & s_{1p} \\ \vdots & \ddots & \vdots \\ s_{1p} & \cdots & s_{pp} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{n}\sum_{j=1}^{n}(x_{j1}-\bar{x}_1)^2 & \cdots & \frac{1}{n}\sum_{j=1}^{n}(x_{j1}-\bar{x}_1)(x_{jp}-\bar{x}_p) \\ \vdots & \ddots & \vdots \\ \frac{1}{n}\sum_{j=1}^{n}(x_{j1}-\bar{x}_1)(x_{jp}-\bar{x}_p) & \cdots & \frac{1}{n}\sum_{j=1}^{n}(x_{jp}-\bar{x}_p)^2 \end{bmatrix}$$

Partitioning: Sample Mean Vector

$$\underset{(p\times 1)}{\bar{\mathbf{x}}} = \begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_q \\ \hdashline \bar{x}_{q+1} \\ \vdots \\ \bar{x}_p \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{x}}^{(1)} \\ \hdashline \bar{\mathbf{x}}^{(2)} \end{bmatrix}$$

Sample mean vectors constructed from obs

$$x^{(1)} = [x_1, \ldots, x_q]'$$

$$x^{(2)} = [x_{q+1}, \ldots, x_p]'$$

Partitioning: CoV Matrix

$$\underset{(p\times p)}{\mathbf{S}_n} = \left[\begin{array}{ccc:ccc} s_{11} & \cdots & s_{1q} & s_{1,q+1} & \cdots & s_{1p} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ s_{q1} & \cdots & s_{qq} & s_{q,q+1} & \cdots & s_{qp} \\ \hdashline s_{q+1,1} & \cdots & s_{q+1,q} & s_{q+1,q+1} & \cdots & s_{q+1,p} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ s_{p1} & \cdots & s_{pq} & s_{p,q+1} & \cdots & s_{pp} \end{array}\right]$$

$$= \begin{array}{cc} & \begin{array}{cc} q & p-q \end{array} \\ \begin{array}{c} q \\ p-q \end{array} & \left[\begin{array}{c:c} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \hdashline \mathbf{S}_{21} & \mathbf{S}_{22} \end{array}\right] \end{array}$$

Obs of these vectors

⇒ sample CoV matrix

$$x^{(1)} \Rightarrow S_{11}$$

$$x^{(2)} \Rightarrow S_{22}$$

$$x^{(1)} \,\&\, x^{(2)} \Rightarrow S_{12} = S_{21}$$

There is some partitioning of sample mean vectors and covariance matrices towards the end of the chapter and in the middle of the chapter. The partitioning seems to result in the same mean/covariance structures, just with some sections identified as the partition. Can you give some insight as to why this is useful? Am I missing something? Why do we partition when we can just take the whole mean/covariance and examine just the entries we are interested in?

Q3: Why Partition?

Matrix: is a billion x a billion

The computer, who has to do the transforming & calculations:

This is… going to take a while…

Q3: Why Partition?

There are some matrices that have a much higher order (>2x2), so it'd be easier to partition it!

You can run into unexpected problems the higher the order is. Computers are technically more powerful nowadays, but our coding applications can still have its limits.

~End~

Any lingering questions?