good partitioned covariance Matrices.pdf, math3086-hongkong-L2 Rand vector good same as above pdf.pdf

Random Vectors and Matrices

A ***random vector*** is a vector whose elements are random variables. Similarly a ***random matrix*** is a matrix whose elements are random variables.

Specifically, let $\mathbf{X} = \{X_{ij}\}$ be an $n \times p$ random matrix. Then the expected value of $\mathbf{X}$, denoted by $E(\mathbf{X})$, is the $n \times p$ matrix of numbers (if they exist)

- The expected value of a random matrix

$$\mathrm{E}(\boldsymbol{X}) = \begin{bmatrix} \mathrm{E}(X_{11}) & \mathrm{E}(X_{12}) & \cdots & \mathrm{E}(X_{1p}) \\ \mathrm{E}(X_{21}) & \mathrm{E}(X_{22}) & \cdots & \mathrm{E}(X_{2p}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}(X_{n1}) & \mathrm{E}(X_{n2}) & \cdots & \mathrm{E}(X_{np}) \end{bmatrix}$$

where, for each element of the matrix,

$$E(X_{ij}) = \begin{cases} \displaystyle\int_{-\infty}^{\infty} x_{ij} f_{ij}(x_{ij})\, dx_{ij} & \text{if } X_{ij} \text{ is a continuous random variable with probability density function } f_{ij}(x_{ij}) \\ \displaystyle\sum_{\text{all } x_{ij}} x_{ij} p_{ij}(x_{ij}) & \text{if } X_{ij} \text{ is a discrete random variable with probability function } p_{ij}(x_{ij}) \end{cases}$$

- $\mathrm{E}(\boldsymbol{X} + \boldsymbol{Y}) = \mathrm{E}(\boldsymbol{X}) + \mathrm{E}(\boldsymbol{Y})$

- $\mathrm{E}(\mathbf{A}\boldsymbol{X}\mathbf{B}) = \mathbf{A}\mathrm{E}(\boldsymbol{X})\mathbf{B}$

Example 2.11 (Computing expected values for discrete random variables) Suppose $p = 2$ and $n = 1$, and consider the random vector $\boldsymbol{X}' = [X_1, X_2]$. Let the discrete random variable X_1 have the following probability function

X_1	-1	0	1
$p_1(X_1)$	0.3	0.3	0.4

Similarly, let the discrete random varibale X_2 have the probability function

X_2	0	1
$p_2(X_2)$	0.8	0.2

Calculate $\mathrm{E}(\boldsymbol{X})$.

Then $E(X_1) = \sum_{\text{all } x_1} x_1 p_1(x_1) = (-1)(.3) + (0)(.3) + (1)(.4) = .1.$

Then $E(X_2) = \sum_{\text{all } x_2} x_2 p_2(x_2) = (0)(.8) + (1)(.2) = .2.$

Thus,

$$E(\mathbf{X}) = \begin{bmatrix} E(X_1) \\ E(X_2) \end{bmatrix} = \begin{bmatrix} .1 \\ .2 \end{bmatrix}$$

Mean Vectors and Covariance Matrices

Suppose $\boldsymbol{X} = [X_1, X_2, \ldots, X_p]$ is a $p \times 1$ random vectors. Then each element of $\boldsymbol{X}$ is a random variables with its own marginal probability distribution.

- The marginal mean $\mu_i = \mathrm{E}(X_i), i = 1, 2, \ldots, p$.
- The marginal variance $\sigma_i^2 = \mathrm{E}(X_i - \mu_i)^2, i = 1, 2, \ldots, p$.

Specifically,

$$\mu_i = \begin{cases} \displaystyle\int_{-\infty}^{\infty} x_i f_i(x_i)\,dx_i & \text{if } X_i \text{ is a continuous random variable with probability density function } f_i(x_i) \\ \displaystyle\sum_{\text{all } x_i} x_i p_i(x_i) & \text{if } X_i \text{ is a discrete random variable with probability function } p_i(x_i) \end{cases}$$

$$\sigma_i^2 = \begin{cases} \displaystyle\int_{-\infty}^{\infty} (x_i - \mu_i)^2 f_i(x_i)\,dx_i & \text{if } X_i \text{ is a continuous random variable with probability density function } f_i(x_i) \\ \displaystyle\sum_{\text{all } x_i} (x_i - \mu_i)^2 p_i(x_i) & \text{if } X_i \text{ is a discrete random variable with probability function } p_i(x_i) \end{cases} \quad (2\text{-}25)$$

It will be convenient in later sections to denote the marginal variances by σ_{ii} rather than the more traditional σ_i^2, and consequently, we shall adopt this notation.

- The behavior of any pair of random variables, such as X_i and X_k, is described by their joint probability function, and a measure of the linear association between them is provided by the covariance

$$\sigma_{ik} = \mathrm{E}(X_i - \mu_i)(X_k - \mu_k)$$

$$\sigma_{ik} = E(X_i - \mu_i)(X_k - \mu_k)$$

$$= \begin{cases} \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x_i - \mu_i)(x_k - \mu_k) f_{ik}(x_i, x_k)\,dx_i\,dx_k & \text{if } X_i, X_k \text{ are continuous random variables with the joint density function } f_{ik}(x_i, x_k) \\ \displaystyle\sum_{\text{all } x_i}\sum_{\text{all } x_k} (x_i - \mu_i)(x_k - \mu_k) p_{ik}(x_i, x_k) & \text{if } X_i, X_k \text{ are discrete random variable with joint probability function } p_{ik}(x_i, x_k) \end{cases} \quad (2\text{-}26)$$

and μ_i and μ_k, $i, k = 1, 2, \ldots, p$, are the marginal means. When $i = k$, the covariance becomes the marginal variance.

The means and covariances of the $p \times 1$ random vector $\mathbf{X}$ can be set out as matrices. The expected value of each element is contained in the vector of means $\boldsymbol{\mu} = E(\mathbf{X})$, and the p variances σ_{ii} and the $p(p-1)/2$ distinct covariances $\sigma_{ik}(i < k)$ are contained in the symmetric variance-covariance matrix $\boldsymbol{\Sigma} = E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'$. Specifically,

$$E(\mathbf{X}) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \boldsymbol{\mu} \tag{2-30}$$

and

$$\begin{aligned}
\boldsymbol{\Sigma} &= E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' \\
&= E\left(\begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_p - \mu_p \end{bmatrix} [X_1 - \mu_1, X_2 - \mu_2, \ldots, X_p - \mu_p]\right) \\
&= E\begin{bmatrix} (X_1 - \mu_1)^2 & (X_1 - \mu_1)(X_2 - \mu_2) & \cdots & (X_1 - \mu_1)(X_p - \mu_p) \\ (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)^2 & \cdots & (X_2 - \mu_2)(X_p - \mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ (X_p - \mu_p)(X_1 - \mu_1) & (X_p - \mu_p)(X_2 - \mu_2) & \cdots & (X_p - \mu_p)^2 \end{bmatrix} \\
&= \begin{bmatrix} E(X_1 - \mu_1)^2 & E(X_1 - \mu_1)(X_2 - \mu_2) & \cdots & E(X_1 - \mu_1)(X_p - \mu_p) \\ E(X_2 - \mu_2)(X_1 - \mu_1) & E(X_2 - \mu_2)^2 & \cdots & E(X_2 - \mu_2)(X_p - \mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ E(X_p - \mu_p)(X_1 - \mu_1) & E(X_p - \mu_p)(X_2 - \mu_2) & \cdots & E(X_p - \mu_p)^2 \end{bmatrix}
\end{aligned}$$

or

$$\boldsymbol{\Sigma} = \operatorname{Cov}(\mathbf{X}) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix} \tag{2-31}$$

- ***Statistical independent*** X_i and X_k if

$$P(X_i \le x_i \text{ and } X_k \le x_k) = P(X_i \le x_i)P(X_k \le x_k)$$

or

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

- ***Mutually statistically independent*** of the p continuous random variables $X_1, X_2, \ldots, X_p$ if

$$f_{1,2,\ldots,p}(x_1, x_2, \ldots, x_p) = f_1(x_1)f_2(x_2)\cdots f_p(x_p)$$

for all p-tuples $(x_1, x_2, \ldots, x_p)$.

- ***linear independent*** of X_i, X_k if

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

Example 2.12 (Computing the covariance matrix) Find the covariance matrix for the two random variables X_1 andX_2 introduced in Example 2.11 when their joint probability function $p_{12}(x_1, x_2)$ is represented by the entries in the body of the following table:

x_1 \ x_2	0	1	$p_1(x_1)$
-1	.24	.06	.3
0	.16	.14	.3
1	.40	.00	.4
$p_2(x_2)$	.8	.2	1

We have already shown that $\mu_1 = E(X_1) = .1$ and $\mu_2 = E(X_2) = .2$. (See Example 2.12.) In addition,

$$\sigma_{11} = E(X_1 - \mu_1)^2 = \sum_{\text{all } x_1} (x_1 - .1)^2 p_1(x_1)$$

$$= (-1 - .1)^2(.3) + (0 - .1)^2(.3) + (1 - .1)^2(.4) = .69$$

$$\sigma_{22} = E(X_2 - \mu_2)^2 = \sum_{\text{all } x_2} (x_2 - .2)^2 p_2(x_2)$$

$$= (0 - .2)^2(.8) + (1 - .2)^2(.2)$$

$$= .16$$

$$\sigma_{12} = E(X_1 - \mu_1)(X_2 - \mu_2) = \sum_{\text{all pairs } (x_1, x_2)} (x_1 - .1)(x_2 - .2) p_{12}(x_1, x_2)$$

$$= (-1 - .1)(0 - .2)(.24) + (-1 - .1)(1 - .2)(.06)$$

$$+ \cdots + (1 - .1)(1 - .2)(.00) = -.08$$

$$\sigma_{21} = E(X_2 - \mu_2)(X_1 - \mu_1) = E(X_1 - \mu_1)(X_2 - \mu_2) = \sigma_{12} = -.08$$

Consequently, with $\mathbf{X}' = [X_1, X_2]$,

$$\boldsymbol{\mu} = E(\mathbf{X}) = \begin{bmatrix} E(X_1) \\ E(X_2) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} .1 \\ .2 \end{bmatrix}$$

and

$$\begin{aligned}
\boldsymbol{\Sigma} &= E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' \\
&= E\begin{bmatrix} (X_1 - \mu_1)^2 & (X_1 - \mu_1)(X_2 - \mu_2) \\ (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)^2 \end{bmatrix} \\
&= \begin{bmatrix} E(X_1 - \mu_1)^2 & E(X_1 - \mu_1)(X_2 - \mu_2) \\ E(X_2 - \mu_2)(X_1 - \mu_1) & E(X_2 - \mu_2)^2 \end{bmatrix} \\
&= \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} .69 & -.08 \\ -.08 & .16 \end{bmatrix}
\end{aligned}$$

■

We note that the computation of means, variances, and covariances for *discrete* random variables involves summation (as in Examples 2.12 and 2.13), while analogous computations for *continuous* random variables involve integration.

Because $\sigma_{ik} = E(X_i - \mu_i)(X_k - \mu_k) = \sigma_{ki}$, it is convenient to write the matrix appearing in (2-31) as

$$\boldsymbol{\Sigma} = E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{bmatrix} \tag{2-32}$$

We shall refer to $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as the *population mean* (vector) and *population variance–covariance* (matrix), respectively.

- ***Population correlation coefficient*** ρ_{ik}

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

The correlation coefficient measures the amount of *linear* association between the random variable X_i and X_k.

Detail:

It is frequently informative to separate the information contained in variances σ_{ii} from that contained in measures of association and, in particular, the measure of association known as the *population correlation coefficient* ρ_{ik}. The correlation coefficient ρ_{ik} is defined in terms of the covariance σ_{ik} and variances σ_{ii} and σ_{kk} as

$$\rho_{ik} = \frac{\sigma_{ik}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{kk}}} \qquad (2\text{-}33)$$

The correlation coefficient measures the amount of *linear* association between the random variables X_i and X_k. (See, for example, [2].)

- ***The population correlation matrix*** $\boldsymbol{\rho}$

Let the population correlation matrix be the $p \times p$ symmetric matrix

$$\boldsymbol{\rho} = \begin{bmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{11}}} & \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{1p}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{pp}}} \\ \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \frac{\sigma_{22}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{pp}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{1p}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{pp}}} & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{pp}}} & \cdots & \frac{\sigma_{pp}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{pp}}} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{12} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1p} & \rho_{2p} & \cdots & 1 \end{bmatrix} \qquad (2\text{-}34)$$

and let the $p \times p$ *standard deviation* matrix be

$$\mathbf{V}^{1/2} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{\sigma_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\sigma_{pp}} \end{bmatrix} \tag{2-35}$$

Then it is easily verified (see Exercise 2.23) that

$$\mathbf{V}^{1/2}\boldsymbol{\rho}\mathbf{V}^{1/2} = \boldsymbol{\Sigma} \tag{2-36}$$

and

$$\boldsymbol{\rho} = (\mathbf{V}^{1/2})^{-1}\boldsymbol{\Sigma}(\mathbf{V}^{1/2})^{-1} \tag{2-37}$$

That is, $\boldsymbol{\Sigma}$ can be obtained from $\mathbf{V}^{1/2}$ and $\boldsymbol{\rho}$, whereas $\boldsymbol{\rho}$ can be obtained from $\boldsymbol{\Sigma}$. Moreover, the expression of these relationships in terms of matrix operations allows the calculations to be conveniently implemented on a computer.

Example 2.14 (Computing the correlation matrix from the covariance matrix) Suppose

$$\boldsymbol{\Sigma} = \begin{bmatrix} 4 & 1 & 2 \\ 1 & 9 & -3 \\ 2 & -3 & 25 \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix}$$

Obtain $\mathbf{V}^{1/2}$ and $\boldsymbol{\rho}$

Here

$$\mathbf{V}^{1/2} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & 0 \\ 0 & \sqrt{\sigma_{22}} & 0 \\ 0 & 0 & \sqrt{\sigma_{33}} \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

and

$$(\mathbf{V}^{1/2})^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{5} \end{bmatrix}$$

Consequently, from (2-37), the correlation matrix $\boldsymbol{\rho}$ is given by

$$(\mathbf{V}^{1/2})^{-1}\boldsymbol{\Sigma}(\mathbf{V}^{1/2})^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{5} \end{bmatrix} \begin{bmatrix} 4 & 1 & 2 \\ 1 & 9 & -3 \\ 2 & -3 & 25 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{5} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \frac{1}{6} & \frac{1}{5} \\ \frac{1}{6} & 1 & -\frac{1}{5} \\ \frac{1}{5} & -\frac{1}{5} & 1 \end{bmatrix}$$

■

Partitioning the Covariance Matrix

- Let

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_q \\ \cdots \\ X_{q+1} \\ \cdots \\ X_p \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}^{(1)} \\ \cdots \\ \boldsymbol{X}^{(2)} \end{bmatrix} \text{ and then } \boldsymbol{\mu} = \mathrm{E}\boldsymbol{X} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_q \\ \cdots \\ \mu_{q+1} \\ \cdots \\ \mu_p \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}^{(1)} \\ \cdots \\ \boldsymbol{\mu}^{(2)} \end{bmatrix}$$

- Define

$$\begin{aligned} & \mathrm{E}(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})' \\ = \; & \mathrm{E}\left[\begin{array}{cc} (\boldsymbol{X}^{(1)}-\boldsymbol{\mu}^{(1)})(\boldsymbol{X}^{(1)}-\boldsymbol{\mu}^{(1)})' & (\boldsymbol{X}^{(1)}-\boldsymbol{\mu}^{(1)})(\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)})' \\ (\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)})(\boldsymbol{X}^{(1)}-\boldsymbol{\mu}^{(1)})' & (\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)})(\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)}) \end{array} \right] \\ = \; & \left[\begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array} \right] \end{aligned}$$

- It is sometimes convenient to use $\text{Cov}(\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)})$ note where

$$\text{Cov}(\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}) = \Sigma_{12} = \Sigma_{21}'$$

  is a matrix containing all of the covariance between a component of $\boldsymbol{X}^{(1)}$ and a component of $\boldsymbol{X}^{(2)}$.

The Mean Vector and Covariance Matrix for Linear Combinations of Random Variables

- The linear combination $\mathbf{c}'\boldsymbol{X} = c_1X_1 + \cdots + c_pX_p$ has

$$\text{mean} = \mathrm{E}(\mathbf{c}'\boldsymbol{X}) = \mathbf{c}'\boldsymbol{\mu}$$

$$\text{variance} = \mathrm{Var}(\mathbf{c}'\boldsymbol{X}) = \mathbf{c}'\Sigma\mathbf{c}$$

  where $\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{X})$ and $\Sigma = \mathrm{Cov}(\boldsymbol{X})$.

- Let C be a matrix, then the linear combinations of $\mathbf{Z} = \mathbf{C}\boldsymbol{X}$ have

$$\boldsymbol{\mu}_{\mathbf{Z}} = \mathrm{E}(\mathbf{Z}) = \mathrm{E}(\mathbf{C}\boldsymbol{X}) = \mathbf{C}\boldsymbol{\mu}_{\mathbf{X}}$$

$$\Sigma_{\mathbf{Z}} = \mathrm{Cov}(\mathbf{Z}) = \mathrm{Cov}(\mathbf{C}\mathbf{X}) = \mathbf{C}\Sigma_{\mathbf{X}}\mathbf{C}'$$

2.1. **Quadratic Form Theorem 1.**

Theorem 1. *If $y \sim N(\mu_y, \Sigma_y)$, then*

$$z = Ay \sim N(\mu_z = A\mu_y;\ \Sigma_z = A\Sigma_y A')$$

where A is a matrix of constants.

2.1.1. *Proof.*

$$\begin{aligned}
E(z) &= E(Ay) = AE(y) = A\mu_y \\
\mathrm{var}(z) &= E[(z - E(z))\,(z - E(z))'] \\
&= E[(Ay - A\mu_y)(Ay - A\mu_y)'] \\
&= E[A(y - \mu_y)(y - \mu_y)'A'] \\
&= AE(y - \mu_y)(y - \mu_y)'A' \\
&= A\Sigma_y A'
\end{aligned}$$

2.1.2. *Example.* Let $Y_1, \ldots, Y_n$ denote a random sample drawn from $N(\mu, \sigma^2)$. Then

$$Y = \begin{pmatrix} Y_1 \\ \cdot \\ \cdot \\ Y_n \end{pmatrix} \sim N \left[\begin{pmatrix} \mu \\ \cdot \\ \cdot \\ \mu \end{pmatrix}, \begin{pmatrix} \sigma^2 & \ldots & 0 \\ \cdot & \sigma^2 & \cdot \\ \cdot & & \\ 0 & & \sigma^2 \end{pmatrix} \right] \tag{4}$$

Now Theorem 1 implies that:

$$\begin{aligned} \bar{Y} &= \frac{1}{n} Y_1 + \cdots + \frac{1}{n} Y_n \\ &= \left(\frac{1}{n}, \ldots, \frac{1}{n} \right) Y = AY \\ &\sim N(\mu, \sigma^2/n) \quad \text{since} \\ &\left(\frac{1}{n}, \ldots, \frac{1}{n} \right) \begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix} = \mu \quad \text{and} \\ &\left(\frac{1}{n}, \ldots, \frac{1}{n} \right) \sigma^2 I \begin{pmatrix} \frac{1}{n} \\ \vdots \\ \frac{1}{n} \end{pmatrix} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \end{aligned} \tag{5}$$

- ***Sample Mean***

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

- ***Sample Covariance Matrix***

$$\begin{aligned} S_n &= \begin{bmatrix} s_{11} & \cdots & s_{1p} \\ \vdots & \ddots & \vdots \\ s_{1p} & \cdots & s_{pp} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{n}\sum_{j=1}^{n}(x_{j1}-\bar{x}_1)^2 & \cdots & \frac{1}{n}\sum_{j=1}^{n}(x_{j1}-\bar{x}_1)(x_{jp}-\bar{x}_p) \\ \vdots & \ddots & \vdots \\ \frac{1}{n}\sum_{j=1}^{n}(x_{j1}-\bar{x}_1)(x_{jp}-\bar{x}_p) & \cdots & \frac{1}{n}\sum_{j=1}^{n}(x_{jp}-\bar{x}_p)^2 \end{bmatrix} \end{aligned}$$

Garcia book 3rd ed page 303

6.1 VECTOR RANDOM VARIABLES

The notion of a random variable is easily generalized to the case where several quantities are of interest. A **vector random variable X** is a function that assigns a vector of real numbers to each outcome ζ in S, the sample space of the random experiment. We use uppercase boldface notation for vector random variables. By convention $\mathbf{X}$ is a column vector (n rows by 1 column), so the vector random variable with components $X_1, X_2, \ldots, X_n$ corresponds to

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = [X_1, X_2, \ldots, X_n]^{\mathrm{T}},$$

where "T" denotes the transpose of a matrix or vector. We will sometimes write $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ to save space and omit the transpose unless dealing with matrices. Possible values of the vector random variable are denoted by $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ where x_i corresponds to the value of X_i.

Page 318

6.3 EXPECTED VALUES OF VECTOR RANDOM VARIABLES

In this section we are interested in the characterization of a vector random variable through the expected values of its components and of functions of its components. We focus on the characterization of a vector random variable through its mean vector and its covariance matrix. We then introduce the joint characteristic function for a vector random variable.

The expected value of a function $g(\mathbf{X}) = g(X_1, \ldots, X_n)$ of a vector random variable $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is given by:

$$E[Z] = \begin{cases} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \ldots, x_n) f_{\mathbf{X}}(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \ldots dx_n & \mathbf{X} \text{ jointly continuous} \\ \sum_{x_1} \cdots \sum_{x_n} g(x_1, x_2, \ldots, x_n) p_{\mathbf{X}}(x_1, x_2, \ldots, x_n) & \mathbf{X} \text{ discrete.} \end{cases} \tag{6.25}$$

An important example is $g(\mathbf{X})$ equal to the sum of functions of $\mathbf{X}$. The procedure leading to Eq. (5.26) and a simple induction argument show that:

$$E[g_1(\mathbf{X}) + g_2(\mathbf{X}) + \cdots + g_n(\mathbf{X})] = E[g_1(\mathbf{X})] + \cdots + E[g_n(\mathbf{X})]. \tag{6.26}$$

Another important example is $g(\mathbf{X})$ equal to the product of n individual functions of the components. If $X_1, \ldots, X_n$ are *independent* random variables, then

$$E[g_1(X_1)g_2(X_2) \ldots g_n(X_n)] = E[g_1(X_1)]E[g_2(X_2)] \ldots E[g_n(X_n)]. \tag{6.27}$$

6.3.1 Mean Vector and Covariance Matrix

The mean, variance, and covariance provide useful information about the distribution of a random variable and are easy to estimate, so we are frequently interested in characterizing multiple random variables in terms of their first and second moments. We now introduce the mean vector and the covariance matrix. We then investigate the mean vector and the covariance matrix of a linear transformation of a random vector.

For $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, the **mean vector** is defined as the column vector of expected values of the components X_k:

$$\mathbf{m_X} = E[\mathbf{X}] = E\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \triangleq \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{bmatrix}. \tag{6.28a}$$

Note that we define the vector of expected values as a column vector. In previous sections we have sometimes written $\mathbf{X}$ as a row vector, but in this section and wherever we deal with matrix transformations, we will represent $\mathbf{X}$ and its expected value as a column vector.

The **correlation matrix** has the second moments of $\mathbf{X}$ as its entries:

$$\mathbf{R_X} = \begin{bmatrix} E[X_1^2] & E[X_1X_2] & \ldots & E[X_1X_n] \\ E[X_2X_1] & E[X_2^2] & \ldots & E[X_2X_n] \\ . & . & \ldots & . \\ E[X_nX_1] & E[X_nX_2] & \ldots & E[X_n^2] \end{bmatrix}. \tag{6.28b}$$

The **covariance matrix** has the second-order central moments as its entries:

$$\mathbf{K_X} = \begin{bmatrix} E[(X_1-m_1)^2] & E[(X_1-m_1)(X_2-m_2)] & \ldots & E[(X_1-m_1)(X_n-m_n)] \\ E[(X_2-m_2)(X_1-m_1)] & E[(X_2-m_2)^2] & \ldots & E[(X_2-m_2)(X_n-m_n)] \\ . & . & \ldots & . \\ E[(X_n-m_n)(X_1-m_1)] & E[(X_n-m_n)(X_2-m_2)] & \ldots & E[(X_n-m_n)^2] \end{bmatrix}. \tag{6.28c}$$

Both $\mathbf{R_X}$ and $\mathbf{K_X}$ are $n \times n$ symmetric matrices. The diagonal elements of $\mathbf{K_X}$ are given by the variances $\text{VAR}[X_k] = E[(X_k - m_k)^2]$ of the elements of $\mathbf{X}$. If these elements are uncorrelated, then $\text{COV}(X_j, X_k) = 0$ for $j \neq k$, and $\mathbf{K_X}$ is a diagonal matrix. If the random variables $X_1, \ldots, X_n$ are independent, then they are uncorrelated and $\mathbf{K_X}$ is diagonal. Finally, if the vector of expected values is $\mathbf{0}$, that is, $m_k = E[X_k] = 0$ for all k, then $\mathbf{R_X} = \mathbf{K_X}$.

Example 5.18 Jointly Gaussian Random Variables

The joint pdf of X and Y, shown in Fig. 5.17, is

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-(x^2-2\rho xy+y^2)/2(1-\rho^2)} \qquad -\infty < x, y < \infty. \tag{5.18}$$

We say that X and Y are jointly Gaussian.[1] Find the marginal pdf's.

The marginal pdf of X is found by integrating $f_{X,Y}(x, y)$ over y:

$$f_X(x) = \frac{e^{-x^2/2(1-\rho^2)}}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} e^{-(y^2-2\rho xy)/2(1-\rho^2)}\, dy.$$

[1]This is an important special case of jointly Gaussian random variables. The general case is discussed in Section 5.9.

We complete the square of the argument of the exponent by adding and subtracting ρ^2x^2, that is, $y^2 - 2\rho xy + \rho^2x^2 - \rho^2x^2 = (y - \rho x)^2 - \rho^2x^2$. Therefore

$$f_X(x) = \frac{e^{-x^2/2(1-\rho^2)}}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} e^{-[(y-\rho x)^2-\rho^2x^2]/2(1-\rho^2)}\, dy$$

$$= \frac{e^{-x^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{e^{-(y-\rho x)^2/2(1-\rho^2)}}{\sqrt{2\pi(1-\rho^2)}}\, dy$$

$$= \frac{e^{-x^2/2}}{\sqrt{2\pi}},$$

where we have noted that the last integral equals one since its integrand is a Gaussian pdf with mean ρx and variance $1 - \rho^2$. The marginal pdf of X is therefore a one-dimensional Gaussian pdf with mean 0 and variance 1. From the symmetry of $f_{X,Y}(x, y)$ in x and y, we conclude that the marginal pdf of Y is also a one-dimensional Gaussian pdf with zero mean and unit variance.

Example 6.6

The random variables X_1, X_2, and X_3 have the joint Gaussian pdf

$$f_{X_1,X_2,X_3}(x_1, x_2, x_3) = \frac{e^{-(x_1^2+x_2^2-\sqrt{2}\,x_1x_2+{}^1\!/_2x_3^2)}}{2\pi\sqrt{\pi}}.$$

Find the marginal pdf of X_1 and X_3. Find the conditional pdf of X_2 given X_1 and X_3.

The marginal pdf for the pair X_1 and X_3 is found by integrating the joint pdf over x_2:

$$f_{X_1,X_3}(x_1, x_3) = \frac{e^{-x_3^2/2}}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{e^{-(x_1^2+x_2^2-\sqrt{2}x_1x_2)}}{2\pi/\sqrt{2}}dx_2.$$

The above integral was carried out in Example 5.18 with $\rho = -1/\sqrt{2}$. By substituting the result of the integration above, we obtain

$$f_{X_1,X_3}(x_1, x_3) = \frac{e^{-x_3^2/2}}{\sqrt{2\pi}}\frac{e^{-x_1^2/2}}{\sqrt{2\pi}}.$$

Therefore X_1 and X_3 are independent zero-mean, unit-variance Gaussian random variables.

The conditional pdf of X_2 given X_1 and X_3 is:

$$f_{X_2}(x_2 \mid x_1, x_3) = \frac{e^{-(x_1^2+x_2^2-\sqrt{2}x_1x_2+{}^1\!/_2x_3^2)}}{2\pi\sqrt{\pi}}\frac{\sqrt{2\pi}\sqrt{2\pi}}{e^{-x_3^2/2}e^{-x_1^2/2}}$$

$$= \frac{e^{-({}^1\!/_2x_1^2+x_2^2-\sqrt{2}x_1x_2)}}{\sqrt{\pi}} = \frac{e^{-(x_2-x_1/\sqrt{2}x_1)^2}}{\sqrt{\pi}}.$$

We conclude that X_2 given and X_3 is a Gaussian random variable with mean $x_1/\sqrt{2}$ and variance 1/2.

Example 6.16

Let $\mathbf{X} = (X_1, X_2, X_3)$ be the jointly Gaussian random vector from Example 6.6. Find $E[\mathbf{X}]$ and $\mathbf{K_X}$.

We rewrite the joint pdf as follows:

$$f_{X_1,X_2,X_3}(x_1, x_2, x_3) = \frac{e^{-(x_1^2+x_2^2-2\frac{1}{\sqrt{2}}x_1x_2)}}{2\pi\sqrt{1-\left(-\frac{1}{\sqrt{2}}\right)^2}}\frac{e^{-x_3^2/2}}{\sqrt{2\pi}}.$$

We see that X_3 is a Gaussian random variable with zero mean and unit variance, and that it is independent of X_1 and X_2. We also see that X_1 and X_2 are jointly Gaussian with zero mean and unit variance, and with correlation coefficient

$$\rho_{X_1X_2} = -\frac{1}{\sqrt{2}} = \frac{\text{COV}(X_1, X_2)}{\sigma_{X_1}\sigma_{X_2}} = \text{COV}(X_1, X_2).$$

Therefore the vector of expected values is: $\mathbf{m_X} = \mathbf{0}$, and

$$\mathbf{K_X} = \begin{bmatrix} 1 & -\frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We now develop compact expressions for $\mathbf{R_X}$ and $\mathbf{K_X}$. If we multiply $\mathbf{X}$, an $n \times 1$ matrix, and $\mathbf{X}^{\mathrm{T}}$, a $1 \times n$ matrix, we obtain the following $n \times n$ matrix:

$$\mathbf{X}\mathbf{X}^{\mathrm{T}} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} [X_1, X_2, \ldots, X_n] = \begin{bmatrix} X_1^2 & X_1X_2 & \ldots & X_1X_n \\ X_2X_1 & X_2^2 & \ldots & X_2X_n \\ . & . & \ldots & . \\ X_nX_1 & X_nX_2 & \ldots & X_n^2 \end{bmatrix}.$$

If we define the expected value of a matrix to be the matrix of expected values of the matrix elements, then we can write the correlation matrix as:

$$\mathbf{R_X} = E[\mathbf{X}\mathbf{X}^{\mathrm{T}}]. \tag{6.29a}$$

The covariance matrix is then:

$$\begin{aligned} \mathbf{K_X} &= E[(\mathbf{X} - \mathbf{m_X})(\mathbf{X} - \mathbf{m_X})^{\mathrm{T}}] \\ &= E[\mathbf{X}\mathbf{X}^{\mathrm{T}}] - \mathbf{m_X}\, E[\mathbf{X}^{\mathrm{T}}] - E[\mathbf{X}]\mathbf{m_X}^{\mathrm{T}} + \mathbf{m_X}\mathbf{m_X}^{\mathrm{T}} \\ &= \mathbf{R_X} - \mathbf{m_X}\mathbf{m_X}^{\mathrm{T}}. \end{aligned} \tag{6.29b}$$

6.3.2 Linear Transformations of Random Vectors

Many engineering systems are linear in the sense that will be elaborated on in Chapter 10. Frequently these systems can be reduced to a linear transformation of a vector of random variables where the "input" is $\mathbf{X}$ and the "output" is $\mathbf{Y}$:

$$\mathbf{Y} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_n \\ a_{21} & a_{22} & \ldots & a_{2n} \\ . & . & \ldots & . \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = \mathbf{AX}.$$

The expected value of the kth component of $\mathbf{Y}$ is the inner product (dot product) of the kth row of $\mathbf{A}$ and $\mathbf{X}$:

$$E[Y_k] = E\left[\sum_{j=1}^{n} a_{kj}X_j\right] = \sum_{j=1}^{n} a_{kj}E[X_j].$$

Each component of $E[\mathbf{Y}]$ is obtained in this manner, so:

$$\begin{aligned} \mathbf{m_Y} = E[\mathbf{Y}] = \begin{bmatrix} \sum_{j=1}^{n} a_{1j}E[X_j] \\ \sum_{j=1}^{n} a_{2j}E[X_j] \\ \vdots \\ \sum_{j=1}^{n} a_{nj}E[X_j] \end{bmatrix} &= \begin{bmatrix} a_{11} & a_{12} & \ldots & a_n \\ a_{21} & a_{22} & \ldots & a_{2n} \\ . & . & \ldots & . \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix} \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{bmatrix} \\ &= \mathbf{A}E[\mathbf{X}] = \mathbf{A}\mathbf{m_X}. \end{aligned} \tag{6.30a}$$

The covariance matrix of $\mathbf{Y}$ is then:

$$\begin{aligned}\mathbf{K_Y} &= E[(\mathbf{Y} - \mathbf{m_Y})(\mathbf{Y} - \mathbf{m_Y})^T] = E[(\mathbf{AX} - \mathbf{Am_X})(\mathbf{AX} - \mathbf{Am_X})^T] \\ &= E[\mathbf{A}(\mathbf{X} - \mathbf{m_X})(\mathbf{X} - \mathbf{m_X})^T\mathbf{A}^T] = \mathbf{A}E[(\mathbf{X} - \mathbf{m_X})(\mathbf{X} - \mathbf{m_X})^T]\mathbf{A}^T \\ &= \mathbf{AK_XA}^T, \end{aligned} \tag{6.30b}$$

where we used the fact that the transpose of a matrix multiplication is the product of the transposed matrices in reverse order: $\{\mathbf{A}(\mathbf{X} - \mathbf{m_X})\}^T = (\mathbf{X} - \mathbf{m_X})^T\mathbf{A}^T$.

The **cross-covariance** matrix of two random vectors $\mathbf{X}$ and $\mathbf{Y}$ is defined as:

$$\mathbf{K_{XY}} = E[(\mathbf{X} - \mathbf{m_X})(\mathbf{Y} - \mathbf{m_Y})^T] = E[\mathbf{XY}^T] - \mathbf{m_X}\mathbf{m_Y}^T = \mathbf{R_{XY}} - \mathbf{m_X}\mathbf{m_Y}^T.$$

We are interested in the cross-covariance between $\mathbf{X}$ and $\mathbf{Y} = \mathbf{AX}$:

$$\begin{aligned}\mathbf{K_{XY}} &= E[\mathbf{X} - \mathbf{m_X})(\mathbf{Y} - \mathbf{m_Y})^T] = E[(\mathbf{X} - \mathbf{m_X})(\mathbf{X} - \mathbf{m_X})^T\mathbf{A}^T] \\ &= \mathbf{K_X A}^T. \end{aligned} \tag{6.30c}$$

Example 6.17 Transformation of Uncorrelated Random Vector

Suppose that the components of $\mathbf{X}$ are uncorrelated and have unit variance, then $\mathbf{K_X} = \mathbf{I}$, the identity matrix. The covariance matrix for $\mathbf{Y} = \mathbf{AX}$ is

$$\mathbf{K_Y} = \mathbf{AK_XA}^T = \mathbf{AIA}^T = \mathbf{AA}^T. \tag{6.31}$$

In general $\mathbf{K_Y} = \mathbf{AA}^T$ is not a diagonal matrix and so the components of $\mathbf{Y}$ are correlated. In Section 6.6 we discuss how to find a matrix $\mathbf{A}$ so that Eq. (6.31) holds for a given $\mathbf{K_Y}$. We can then generate a random vector $\mathbf{Y}$ with any desired covariance matrix $\mathbf{K_Y}$.

Suppose that the components of $\mathbf{X}$ are correlated so $\mathbf{K_X}$ is not a diagonal matrix. In many situations we are interested in finding a transformation matrix $\mathbf{A}$ so that $\mathbf{Y} = \mathbf{AX}$ has uncorrelated components. This requires finding $\mathbf{A}$ so that $\mathbf{K_Y} = \mathbf{AK_XA}^T$ is a diagonal matrix. In the last part of this section we show how to find such a matrix $\mathbf{A}$.

Example 6.18 Transformation to Uncorrelated Random Vector

Suppose the random vector X_1, X_2, and X_3 in Example 6.16 is transformed using the matrix:

$$\mathbf{A} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Find the $E[\mathbf{Y}]$ and $\mathbf{K_Y}$.

Since $\mathbf{m_X} = \mathbf{0}$, then $E[\mathbf{Y}] = \mathbf{Am_X} = \mathbf{0}$. The covariance matrix of $\mathbf{Y}$ is:

$$\mathbf{K_Y} = \mathbf{AK_XA}^{\mathrm{T}} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & -\frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \frac{1}{2}\begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 - \frac{1}{\sqrt{2}} & 1 + \frac{1}{\sqrt{2}} & 0 \\ 1 - \frac{1}{\sqrt{2}} & -\left(1 + \frac{1}{\sqrt{2}}\right) & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 - \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 1 + \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The linear transformation has produced a vector of random variables $\mathbf{Y} = (Y_1, Y_2, Y_3)$ with components that are uncorrelated.