Ch2. Random Vectors and Matrices

1 Definition

Definition 1.1 *A* **random vector** *or* **random matrix** *is a vector or matrix whose elements are random variables.*

In formally, a **random variable** is defined as a variable whose value depends on the outcome of a chance experiment. (Formally, a random variable is a function defined for each element of a sample space.)

In terms of experimental structure, we have two kinds of random vectors:

1. A vector containing a measurement on each of n different individuals or experimental units. When the same variable is observed on each of n units selected at random, the n random variables $y_1, y_2, \cdots, y_n$ in the vector are typically uncorrelated and have the same variance.

2. A vector consisting of p different measurements on one individual or experimental unit. The p random variables thus obtained are typically correlated and have different variables.

2 Means, Variances, Covariances, and Correlations (univariate)

Let $f(y)$ denote the density of the random variable y, the mean or expected value of y is defined as

$$\mu = E(y) = \int_{-\infty}^{\infty} y f(y) dy.$$

In general, for a function $u(y)$, we have

$$E[u(y)] = \int_{-\infty}^{\infty} u(y) f(y) dy.$$

A variance of a random variable y is defined as

$$\sigma^2 = var(y) = E(y - \mu)^2.$$

A square root of the variance is called the standard deviation,

$$\sigma = \sqrt{var(y)} = \sqrt{E(y-\mu)^2}.$$

For any two variables y_i and y_j, the covariance is

$$\sigma_{ij} = cov(y_i, y_j) = E[(y_i - \mu_i)(y_j - \mu_j)].$$

To standardize σ_{ij}, we divide it by (the product of) the standard deviations of y_i and y_j to obtain the correlation

$$\rho_{ij} = corr(y_i, y_j) = \frac{\sigma_{ij}}{\sigma_i \sigma_j}.$$

The random variable y_i and y_j are said to be independent if their joint density factors into the product of their marginal densities:

$$f(y_i, y_j) = f_i(y_i) f_j(y_j),$$

where the marginal density $f_i(y_i)$ is defined as

$$f_i(y_i) = \int_{-\infty}^{\infty} f(y_i, y_j) dy_j.$$

The independence implies the following two properties:

1. $E(y_i y_j) = E(y_i)E(y_j)$ if y_i and y_j are independent.

2. $\sigma_{ij} = cov(y_i, y_j) = 0$ if y_i and y_j are independent.

Note that the converse of the second property is not true in general; that is, $\sigma_{ij} = 0$ does not imply independence. It is only true in the case that y_i and y_j have a bivariate normal distribution.

3 Mean Vectors and Covariance Matrices for Random Vectors

3.1 Mean vector

$$E(\boldsymbol{y}) = E\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix} = \begin{pmatrix} E(y_1) \\ E(y_2) \\ \vdots \\ E(y_p) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix} = \boldsymbol{\mu},$$

where $E(y_i) = \mu_i$.

For the mean vectors, we have

$$E(\boldsymbol{x} + \boldsymbol{y}) = E(\boldsymbol{x}) + E(\boldsymbol{y}).$$

By analogy with $E(\boldsymbol{y})$, we define the expected value of a random matrix $\boldsymbol{Z}$ as the matrix of expected values:

$$E(\boldsymbol{Z}) = E \begin{pmatrix} z_{11} & z_{12} & \cdots & z_{1p} \\ z_{21} & z_{22} & \cdots & z_{2p} \\ \vdots & \vdots & & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{np} \end{pmatrix} = \begin{pmatrix} E(z_{11}) & E(z_{12}) & \cdots & E(z_{1p}) \\ E(z_{21}) & E(z_{22}) & \cdots & E(z_{2p}) \\ \vdots & \vdots & & \vdots \\ E(z_{n1}) & E(z_{n2}) & \cdots & E(z_{np}) \end{pmatrix}$$

3.2 Covariance Matrix

$$E[(\boldsymbol{y} - \boldsymbol{\mu})(\boldsymbol{y} - \boldsymbol{\mu})'] = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix} = \Sigma,$$

where $\sigma_{ii} = var(y_i)$ and $\sigma_{ij} = cov(y_i, y_j)$.

3.3 Generalized Variance

A measure of overall variability in the population of $\boldsymbol{y}$'s can be defined as the determinant of Σ:

$$Generalized \quad variance = |\Sigma|.$$

If $|\Sigma|$ is small, the $\boldsymbol{y}$'s are concentrated closer to $\boldsymbol{\mu}$ than of $|\Sigma|$ is large. A small value of $|\Sigma|$ may also indicate that the variables $y_1, y_2, \cdots, y_p$ in $\boldsymbol{y}$ are highly intercorrelated, in which case the $\boldsymbol{y}$'s tend to occupy a subspace of the p dimensions.

3.4 Standardized Distance

To obtain a useful measure of distance between $\boldsymbol{y}$ and $\boldsymbol{\mu}$, we need to take into account the variance and covariance of the y_i's in $\boldsymbol{y}$. The standardized distance is defined as

$$Standardized \quad distance = (\boldsymbol{y} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{y} - \boldsymbol{\mu}).$$

This distance is often called a Mahalanobis distance.

3.5 Correlation Matrices

The correlation matrix is defined as

$$\boldsymbol{P}_\rho = (\rho_{ij}) = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{pmatrix},$$

where $\rho_{ij} = \sigma_{ij}/[\sigma_i \sigma_j]$. If we define

$$\boldsymbol{D}_\sigma = diag(\sigma_1, \sigma_2, \cdots, \sigma_p),$$

we have

$$\boldsymbol{P}_\rho = \boldsymbol{D}_\sigma^{-1} \Sigma \boldsymbol{D}_\sigma^{-1},$$

and

$$\Sigma = \boldsymbol{D}_\sigma \boldsymbol{P}_\sigma \boldsymbol{D}_\sigma.$$

4 Linear Functions of Random Vectors

4.1 Means

Theorem 4.1 *If $\boldsymbol{a}$ is a $p \times 1$ vector of constants and $\boldsymbol{y}$ is a $p \times 1$ random vector with mean vector $\boldsymbol{\mu}$, then the mean of $z = \boldsymbol{a}'\boldsymbol{y}$ is given by*

$$\mu_z = E(\boldsymbol{a}'\boldsymbol{y}) = \boldsymbol{a}'E(\boldsymbol{y}) = \boldsymbol{a}'\boldsymbol{\mu}.$$

Theorem 4.2 *Suppose $\boldsymbol{y}$ is a random vector, $\boldsymbol{X}$ is a random matrix, $\boldsymbol{a}$ and $\boldsymbol{b}$ are vectors of constants, and $\boldsymbol{A}$ and $\boldsymbol{B}$ are matrices of constants. Then, assuming the matrices and the vectors in each product are conformable, we have the following expected values:*

(a) $E(\boldsymbol{A}\boldsymbol{y}) = \boldsymbol{A}E(\boldsymbol{y})$.

(b) $E(\boldsymbol{a}'\boldsymbol{X}\boldsymbol{b}) = \boldsymbol{a}'E(\boldsymbol{X})\boldsymbol{b}$,

(c) $E(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}) = \boldsymbol{A}E(\boldsymbol{X})\boldsymbol{B}$.

4.2 Variances and Covariances

Theorem 4.3 *If $\boldsymbol{a}$ is a $p \times 1$ vector of constants and $\boldsymbol{y}$ is a $p \times 1$ random vector with covariance matrix Σ, then the variance of $z = \boldsymbol{a}'\boldsymbol{y}$ is given by*

$$\sigma_z^2 = var(\boldsymbol{a}'\boldsymbol{y}) = \boldsymbol{a}'\Sigma\boldsymbol{a}.$$

Corollary 4.1 *If $\boldsymbol{a}$ and $\boldsymbol{b}$ are $p \times 1$ vectors of constants, then*

$$cov(\boldsymbol{a}'\boldsymbol{y}, \boldsymbol{b}'\boldsymbol{y}) = \boldsymbol{a}'\Sigma\boldsymbol{b}.$$

Theorem 4.4 *Let $\boldsymbol{z} = \boldsymbol{A}\boldsymbol{y}$ and $\boldsymbol{w} = \boldsymbol{B}\boldsymbol{y}$, where $\boldsymbol{A}$ us a $k \times p$ matrix of constants, $\boldsymbol{B}$ is an $m \times p$ matrix of constants, and $\boldsymbol{y}$ is a $p \times 1$ random vector with covariance matrix Σ. Then*

(a) $cov(\boldsymbol{z}) = cov(\boldsymbol{A}\boldsymbol{y}) = \boldsymbol{A}\Sigma\boldsymbol{A}'$,

(b) $cov(\boldsymbol{z}, \boldsymbol{w}) = cov(\boldsymbol{A}\boldsymbol{y}, \boldsymbol{B}\boldsymbol{y}) = \boldsymbol{A}\Sigma\boldsymbol{B}'$.