Review of Matrix Algebra

* In this section I will give a brief review of some key points in matrix algebra.

* This is not exhaustive and students are required to have taken a course in linear algebra to the level of Math 2R03.

* Many basic results are given in the appendix to Chapter 2 of the textbook.

Some Notation

* I will denote vector or matrix quantities in bold.
* All vectors are considered to be column vectors (i.e. a vector is a matrix with only one column).
* I will use $\boldsymbol{A}^t$ to denote the transpose of a vector or matrix.
* $\boldsymbol{I}_p$ will be the $p \times p$ identity matrix.
* $\mathbf{0}_p$ and $\mathbf{1}_p$ will be vectors of length p with all entries equal to 0 or 1 respectively

Some Basics

* If c is a constant then $c\boldsymbol{A}$ multiplies each element of $\boldsymbol{A}$ by c.

* If $\boldsymbol{A}$ and $\boldsymbol{B}$ are of the same size then $\boldsymbol{A}+\boldsymbol{B}$ is the element-wise sum of the two matrices.

* If $\boldsymbol{A}$ is of dimension $n \times m$ and $\boldsymbol{B}$ is of dimension $m \times p$ then $\boldsymbol{C} = \boldsymbol{AB}$ is a $n \times p$ matrix with

$$C_{ij} = \sum_{k=1}^{m} A_{ik} B_{kj}.$$

* It is important to note that matrix multiplication is not commutative ($\boldsymbol{AB} \neq \boldsymbol{BA}$).

Properties of the Determinant

* The determinant of a 2×2 matrix is

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

* For larger square matrices the determinant is found through a recursive definition.

* The determinant of a diagonal matrix is the product of the diagonal elements.

Properties of the Determinant

* If a single column (or row) of $\boldsymbol{A}$ is multiplied by a constant c then the determinant is also multiplied by that constant.

* If a multiple of one column (or row) is added to another then the determinant is unchanged.

* If two columns (or rows) are interchanged then the sign of the determinant changes.

* If any column is a multiple of another then the determinant is 0.

* If $\boldsymbol{A}$ and $\boldsymbol{B}$ are both $p \times p$ matrices then $|\boldsymbol{AB}| = |\boldsymbol{BA}| = |\boldsymbol{A}||\boldsymbol{B}|$

The Inverse

* If $\boldsymbol{A}$ is a $p \times p$ matrix then the Inverse of $\boldsymbol{A}$ is the $p \times p$ matrix $\boldsymbol{A}^{-1}$ such that

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_p$$

* A square matrix $\boldsymbol{A}$ has an inverse if, and only if, it is non-singular so $|\boldsymbol{A}| \neq 0$.

* The inverse of a diagonal matrix is a diagonal matrix with elements equal to the reciprocal of the original matrix.

* $\left(\boldsymbol{A}^t\right)^{-1} = \left(\boldsymbol{A}^{-1}\right)^t$.

* If A and B are square matrices then $(\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$.

The Trace

* The trace of a square matrix is the sum of the diagonal

$$\mathrm{tr}(\boldsymbol{A}) \;=\; \sum_i A_{ii}$$

* A useful result is that if $\boldsymbol{AB}$ is a square matrix then

$$\mathrm{tr}(\boldsymbol{AB}) \;=\; \mathrm{tr}(\boldsymbol{BA})$$

Important Square Matrices

* A square matrix $\boldsymbol{A}$ is symmetric if, and only if, $\boldsymbol{A}^t = \boldsymbol{A}$.

* A square matrix $\boldsymbol{A}$ is orthogonal if, and only if,

$$\boldsymbol{A}\boldsymbol{A}^t = \boldsymbol{A}^t\boldsymbol{A} = \boldsymbol{I}.$$

* A square matrix $\boldsymbol{X}$ is idempotent if, and only if,

$$\boldsymbol{A}^2 = \boldsymbol{A}\boldsymbol{A} = \boldsymbol{A}.$$

* A useful property of an idempotent matrix is that its trace is always an integer and is equal to therank of the matrix.

Eigenvalues and Eigenvectors

* For any square matrix $\boldsymbol{A}$, an eigenvalue, λ and corresponding eigenvector, $\boldsymbol{v} \neq \mathbf{0}$, satisfy

$$\boldsymbol{A}\boldsymbol{e} \;=\; \lambda \boldsymbol{e}$$

* The eigenvalues can be found as the p solutions to the equation

$$|\boldsymbol{A} - \lambda \boldsymbol{I}_p| \;=\; 0$$

* The eigenvector corresponding to an eigenvalue is not uniquely determined but will conventionally be chosen such that they have length one and are orthogonal to each other.

* The Spectral Decomposition of a $p \times p$ matrix $\boldsymbol{A}$ is given by

$$\boldsymbol{A} \;=\; \sum_{i=1}^{p} \lambda_i \boldsymbol{e}_i \boldsymbol{e}_i^t$$

Quadratic Forms and Positive Definite Matrices

* Suppose that $\boldsymbol{x}$ is a vector of length p and $\boldsymbol{A}$ is a symmetric $p \times p$ matrix, then $\boldsymbol{x}^t\boldsymbol{A}\boldsymbol{x}$ is called a Quadratic Form.

* If $\boldsymbol{A}$ satisfies that

$$\boldsymbol{x}^t\boldsymbol{A}\boldsymbol{x} \;\geqslant\; 0 \qquad \text{for every } p\text{-vector } \boldsymbol{x}$$

then $\boldsymbol{A}$ is said to be nonnegative definite.

* If $\boldsymbol{A}$ is nonnegative definite and

$$\boldsymbol{x}^t\boldsymbol{A}\boldsymbol{x} \;=\; 0 \quad \Longleftrightarrow \quad \boldsymbol{x} = \boldsymbol{0}$$

then $\boldsymbol{A}$ is said to be positive definite.

Square Root Matrix

* In many situations it can be useful to have a square root matrix, for a positive definite matrix, satisfying

$$\boldsymbol{A} = \boldsymbol{A}^{1/2}\boldsymbol{A}^{1/2}$$

* Suppose that the spectral decomposition of $\boldsymbol{A}$ is

$$\boldsymbol{A} = \sum_{i=1}^{p} \lambda_1 \boldsymbol{e}_i \boldsymbol{e}_i^t = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}$$

where $\boldsymbol{P}$ is the orthogonal matrix of eigenvectors and $\boldsymbol{\Lambda}$ is a diagonal matrix with the eigenvalues on the diagonal. Then we can see that

$$\boldsymbol{A} = \left(\boldsymbol{P}\boldsymbol{\Lambda}^{1/2}\boldsymbol{P}^t\right)\left(\boldsymbol{P}\boldsymbol{\Lambda}^{1/2}\boldsymbol{P}^t\right)$$

* Also we have

$$\boldsymbol{A}^{-1} = \left(\boldsymbol{P}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{P}^t\right)\left(\boldsymbol{P}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{P}^t\right)$$

Random Vectors and Matrices

* Throughout this course we shall be dealing with random vectors and matrices.

* A random vector (matrix) is one whose elements are random variables in the usual univariate sense.

* The Expected Value of a random matrix $\boldsymbol{X} = \{X_{ij}\}$ is

$$\mathsf{E}[\boldsymbol{X}] = \begin{pmatrix} \mathsf{E}[X_{11}] & \mathsf{E}[X_{12}] & \cdots & \mathsf{E}[X_{1p}] \\ \mathsf{E}[X_{21}] & \mathsf{E}[X_{22}] & \cdots & \mathsf{E}[X_{2p}] \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{E}[X_{n1}] & \mathsf{E}[X_{n2}] & \cdots & \mathsf{E}[X_{np}] \end{pmatrix}$$

* If $\boldsymbol{X}$ is a random vector and $\boldsymbol{A}$ and $\boldsymbol{B}$ are conformable matrices of constants then

$$\mathsf{E}[\boldsymbol{AXB}] = \boldsymbol{A}\mathsf{E}[\boldsymbol{X}]\boldsymbol{B}.$$

Joint Probability Distributions

* Suppose that $\boldsymbol{X}$ is a random vector, we can describe its distribution with a joint probability density function $f_{\boldsymbol{X}}(x_1, \ldots, x_p)$.

* A joint pdf satisfies the considitions
  1. $f_{\boldsymbol{X}}(x_1, \ldots, x_p) \geqslant 0$ for all $(x_1, \ldots, x_p) \in \mathbb{R}^p$.
  2. $$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\boldsymbol{X}}(x_1, \ldots, x_p)\, dx_1 \cdots dx_p = 1$$

* If any components of $\boldsymbol{X}$ are discrete then the corresponding integral becomes a sum over the possible values.

* Note that, for many distributions, the support

$$\mathcal{X} = \{(x_1, \ldots, x_p) \mid f_{\boldsymbol{X}}(x_1, \ldots, x_p) > 0\}$$

is not equal to $\mathbb{R}^p$ so care must be taken in the limits of the integrals.

Marginal and Conditional Distributions

* To find the marginal pdf for X_i we integrate over all of the other components

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\boldsymbol{X}}(x_1, \ldots, x_p)\, dx_1 \cdots dx_{i-1}\, dx_{i+1} \cdots dx_p$$

* We can also find the marginal joint pdf for some components of some components in a similar way.

* The full conditional distribution of X_i given values for the other components is

$$f_{X_i|\boldsymbol{X}_{-i}}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_p) = \frac{f_{\boldsymbol{X}}(x_1, \ldots, x_p)}{\int_{-\infty}^{\infty} f_{\boldsymbol{X}}(x_1, \ldots, x_p)\, dx_i}$$

Expectation

* Suppose that g is a scalar function on $\mathbb{R}^p$. Then the expectation of $g(\boldsymbol{X})$ can be found from the joint probability density

$$\mathsf{E}[g(\boldsymbol{X})] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \ldots, x_p) f_{\boldsymbol{X}}(x_1, \ldots, x_p)\, dx_1 \cdots dx_p$$

* We can, for example, find the marginal mean of X_i by taking $g(x_1, \ldots, x_p) = x_i$ so

$$\begin{aligned}
\mathsf{E}[X_1] &= \mathsf{E}[g(\boldsymbol{X})] \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \ldots, x_p) f_{\boldsymbol{X}}(x_1, \ldots, x_p)\, dx_1 \cdots dx_p \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_i f_{\boldsymbol{X}}(x_1, \ldots, x_p)\, dx_1 \cdots dx_p \\
&= \int_{-\infty}^{\infty} x_i f_{X_i}(x_i)\, dx_i
\end{aligned}$$

Mean Vectors and Covariance Matrices

Definition 1

Suppose that $\boldsymbol{X} = (X_1, \ldots, X_p)^t$ *is a random vector. Then the mean of* $\boldsymbol{X}$ *is*

$$\mathsf{E}[\boldsymbol{X}] = \boldsymbol{\mu} = (\mathsf{E}[X_1], \mathsf{E}[X_2], \ldots, \mathsf{E}[X_p])^t$$

and the covariance matrix of $\boldsymbol{X}$ *is the matrix*

$$\mathrm{Var}(\boldsymbol{X}) = \boldsymbol{\Sigma} = \mathsf{E}\left[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^t\right]$$

$$= \begin{pmatrix} \mathsf{E}\left[(X_1 - \mu_1)^2\right] & \mathsf{E}\left[(X_1 - \mu_1)(X_2 - \mu_2)\right] & \cdots & \mathsf{E}\left[(X_1 - \mu_1)(X_p - \mu_p)\right] \\ \mathsf{E}\left[(X_1 - \mu_1)(X_2 - \mu_2)\right] & \mathsf{E}\left[(X_2 - \mu_2)^2\right] & \cdots & \mathsf{E}\left[(X_2 - \mu_2)(X_p - \mu_p)\right] \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{E}\left[(X_1 - \mu_1)(X_p - \mu_p)\right] & \mathsf{E}\left[(X_2 - \mu_2)(X_p - \mu_p)\right] & \cdots & \mathsf{E}\left[(X_p - \mu_p)^2\right] \end{pmatrix}$$

Correlation Matrix

* The population correlation between 2 random variables X_i and X_j is

$$\rho_{ij} = \frac{\mathsf{E}\left[(X_i - \mu_i)(X_j - \mu_j)\right]}{\sqrt{\mathsf{Var}(X_i)}\sqrt{\mathsf{Var}(X_j)}} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}$$

* The correlation matrix for the random vector $\boldsymbol{X} = (X_1, \ldots, X_p)^t$ is then

$$\boldsymbol{\rho} = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{12} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1p} & \rho_{2p} & \cdots & 1 \end{pmatrix}$$

* If $\boldsymbol{V}$ is a diagonal matrix with elements σ_{ii} then we have the relationship

$$\boldsymbol{\Sigma} = \boldsymbol{V}^{1/2}\boldsymbol{\rho}\boldsymbol{V}^{1/2}$$

Independence in $\mathbb{R}^2$

* Two random variable X_1 and X_2 are independent if, and only if,

$$f_{X_1,X_2}(x_1,x_2) \;=\; f_{X_1}(x_1)f_{X_2}(x_2)$$

for every point $(x_1,x_2) \in \mathbb{R}^2$.

* A consequence of this is then that

$$\mathsf{E}[g_1(X_1)g_2(X_2)] \;=\; \mathsf{E}[g_1(X_1)]\mathsf{E}[g_2(X_2)]$$

* Hence we see that if X_1 and X_2 are independent then $\mathsf{Cov}(X_1,X_2) = 0$.

* The inverse is not true in general.

Independence in $\mathbb{R}^p$

* The random variables $X_1, \ldots, X_p$ are mutually independent if, and only if,

$$f_{X_{i_1},\ldots,X_{i_k}}(x_{i_1}, \ldots, x_{i_k}) \;=\; f_{X_{i_1}}(x_{i_1}) \cdots f_{X_{i_k}}(x_{i_k})$$

for every set of indices $\{i_1, \ldots, i_k\} \subset \{1, \ldots, p\}$.

* A consequence is that

$$f_{\boldsymbol{X}}(x_1, \ldots, x_p) \;=\; f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_p}(x_p)$$

* The random variables $X_1, \ldots, X_p$ are pairwise independent if, and only if,

$$f_{X_i,X_j}(x_i, \ldots, x_j) \;=\; f_{X_i}(x_i) \cdots f_{X_j}(x_j)$$

for every pair of indices $i \neq j$.

Partitioned Random Vectors

* It will sometimes be useful to partition the random vector $\boldsymbol{X}$ as

$$\boldsymbol{X} = \left(\begin{array}{c} X_1 \\ \vdots \\ X_q \\ \hline X_{q+1} \\ \vdots \\ X_p \end{array} \right) = \left(\begin{array}{c} \boldsymbol{X}^{(1)} \\ \hline \boldsymbol{X}^{(2)} \end{array} \right)$$

* This induces partitioning of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as

$$\boldsymbol{\mu} = \left(\begin{array}{c} \boldsymbol{\mu}^{(1)} \\ \hline \boldsymbol{\mu}^{(2)} \end{array} \right) \qquad \boldsymbol{\Sigma} = \left(\begin{array}{c|c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array} \right)$$

* $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}^t$. is the matrix of covariances between a component of $\boldsymbol{X}^{(1)}$ and a component of $\boldsymbol{X}^{(2)}$.

Linear Combinations

* Suppose that $\boldsymbol{c} = (c_1, \ldots, c_p)^t$ is a vector of constants then

$$Y = \boldsymbol{c}^t \boldsymbol{X} = \sum_{i=1}^{p} c_i X_i$$

  is a linear combination.

* Y is a scalar random variable.

* The mean and variance of Y are

$$\mathsf{E}[Y] = \boldsymbol{c}^t \boldsymbol{\mu} \qquad \mathsf{Var}(Y) = \boldsymbol{c}^t \boldsymbol{\Sigma} \boldsymbol{c}$$