Mathematics for Biostatistics

Charles A. Rohde

2011

Contents

1 Numbers, Sets and Functions

1.1 Introduction

Biological systems and problems arising in medicine and public health are often complex. To be understood quantitatively we reduce complexity by making simplifying assumptions. This replaces the system under investigation by an **imaginary model** which is simple enough to be described mathematically and verified experimentally. The results of any mathematical and statistical analysis will be applicable to the imaginary model. The results will be applicable to the original biological system only to the extent that the assumptions are reasonable.

Because modern public health and medical research focuses on quantitative relationships we use mathematical and statistical models to develop hypotheses subject to experimental verification. Such models ought not to be taken too lightly or too seriously, they are simply inevitable tools of modern research and not too modern at that:

> When Issac Newton communicated the manuscript of his Methodus fluxionum to his friends in 1669 he furnished science with its most powerful and subtle instrument of research.

J.W. Mellor (1931) Higher Mathematics, Longmans, Greena and Company

Since mathematical work is inevitable we need to be aware of some basic principles:

(1) Mathematical derivations of any use take much **time** and **work** and **thought**.

> The first thing to be attended to in reading any algebraic treatise is the gaining of a perfect understanding of the different processes there exhibited, and of their connection with each other. This cannot be attained by a mere reading of the book, however great the attention which may have been given. It is impossible in a mathematical work to fill up every process in the manner in which it must be filled up in the mind of the student before he can be said to have completely mastered it. Many results must be given of which the details are suppressed, such as the additions, multiplications, extractions of square roots, etc. with which the investigations abound. These must not be taken in trust by the student, but must be worked by his own pen, which must not be out of his hand, while engaged in any algebraical process.

DeMorgan (1831) On the Study and Difficulties of Mathematics

(2) Mathematical work is either correct or incorrect which implies that we must constantly ask whether the result makes sense.

(3) Terms must be defined clearly and precisely.

(4) Symbols should be selected to add to clarity, not to detract from it.

(5) Complicated expressions which occur frequently should be replaced by a single symbol.

(6) Be consistent in the use of symbols.

(7) Check all equations for misprints, etc. **caveat emptor !**

1.2 Numbers and Dimensions

Since modelling involves quantitative relationships we note several general properties of numbers. Numbers can be described in two general ways, quantitatively or qualitatively.

Quantitative: Quantitative descriptions of numbers include the actual numerical value and its classification into either a discrete or continuous type.

- The discrete classification includes the *nominal* numbers in which the number merely serves as a label and the *ordinal* numbers which define an ordered label. Nominal number are not subject to the usual arithmetic operations such as addition, subtraction, multiplication and division and there is no ordering. Ordinal numbers may be ordered but no other algebraic operations are permitted.
- Continuous numbers include *interval* and *ratio* numbers. Interval numbers may be added or subtracted as well as ordered but not multiplied or divided. Ratio numbers are subject to all arithmetic operations and ordering.

Qualitative: Qualitative descriptions of numbers specify their dimension. Analysis of the dimensions of numbers (called dimensional analysis) provides an extraordinarily simple, yet powerful, method of checking the correctness of models. Any equation in a model must be dimensionally correct. That is, it must satisfy the following conditions:

(1) Quantities added and subtracted and the result of such operations must have the same dimensions.

(2) Quantities equal to each other must have the same dimension.

(3) Quantities may be multiplied or divided without regard to dimension but the result must have dimension satisfying (2).

(4) Dimensions of a qunatity are independent of its magnitude.

(5) Numbers such as π, e and probabilities do not have a dimension.

(6) Dimensions of a quantity are not changed by multiplication by a dimensionaless number.

The two properties of numbers are related by the folowing two principles:

(1) The units of measurement must not change in any set of calculations

(2) In empirically derived equations (e.g. regression equations) the coefficients implicitly have units equal to

$$\frac{\text{response dimension}}{\text{covariate dimension}}$$

1.3 Real Numbers

The different types of numbers are as follows:

(1) The *natural numbers* $\mathbf{N} = \{1, 2, \ldots\}$

(2) The *integers* $\mathbf{I} = \{\ldots, -2, -1, 0, +1, +2, \ldots\}$

(3) The *rational numbers* $\mathcal{R} = \{x \ : \ x = p/q \ , \ p, q \ \in \ \mathbf{I} \ , \ q \neq 0\}$

(4) The *irrational numbers* needed to solve certain equations e.g. $(x^2 = 2)$. Includes numbers such as $\sqrt{2}, e, \pi, \log(3)$

(5) The *real numbers* $\mathbf{R}$ which include the rational numbers and the irrational numbers.

(6) The *complex numbers* needed to solve equations such as $z^2 = -1$. They are used in statistics in the analysis of time series data. They are discussed in the appendix.

1.4 Indices and Summation Notation

Given an ordered collection of numbers

$$\mathcal{A} = (a_1, a_2, a_3, \ldots, a_n)$$

we write the sum

$$a_1 + a_2 + \cdots + a_n$$

as

$$\sum_{i=1}^{n} a_i$$

The i is an index and is merely an indicator of which elements in $\mathcal{A}$ to be included in the sum. Thus

$$\sum_{i=1}^{n} a_i = \sum_{j=1}^{n} a_j = \sum_{w=1}^{n} a_w$$

A similar notation holds for products e.g.

$$\prod_{i=1}^{n} a_i = \prod_{j=1}^{n} a_j = a_1 \times a_2 \times \cdots \times a_n$$

1.5 Sets

1.5.1 Definitions

Definition 1.1 A set is a collection of points or elements.

(1) The empty set $\emptyset$, is the set containing no points.

(2) All sets under consideration are assumed to consist of points of a fixed non-empty set Ω (called a space).

(3) Points of Ω are denoted by ω or x

(4) Capital letters such as $E_1, E_2, \ldots$ denote sets and $\{\omega\}$ denotes the set consisting of the single point ω.

(5) If ω is a point in the set E we write $\omega \in E$ while if ω is not a point in the set E we write $\omega \notin E$.

(6) To describe a set E we write

$$E = \{\omega : \; S(\omega)\}$$

i.e. E is the set of all points such that the statement $S(\omega)$ is true. Alternatively we shall write $\{\ldots\ldots\}$ where all points in E are written down inside the brackets.

Definition 1.2 A set of sets is called a *class*. Classes are denoted by script letters e.g. $\mathcal{W}$. The set of all subsets of Ω is called the *power set* of Ω and is denoted by 2^{Ω}.

Classes of sets play a fundamental role in mathematics e.g. the open sets of real numbers, the closed sets of real numbers, the Borel sets of real numbers, the σ-algebras used in probability theory.

Definition 1.3 (*set inclusion*) A set E is said to be contained in a set F if $\omega \in E$ implies $\omega \in F$. This relation is written $E \subset F$.

Note that $\emptyset \subset E \subset \Omega$ for every set E and that the relation of set inclusion is reflexive and transitive i.e.

$$E \subset E \; ; \; E \subset F, \; F \subset G \Rightarrow E \subset G$$

Definition 1.4 (*set equality*) Sets E and F are said to be equal if $E \subset F$ and $F \subset E$.

Note that set equality is reflexive, symmetric and transitive i.e.

$$E = E, \; E = F \Rightarrow F = E \;\text{ and }\; E = F, \; F = G \Rightarrow E = G$$

Definition 1.5 (*difference of two sets*) The *difference* of two sets $E - F$ is the set defined as

$$E - F = \{\omega : \; \omega \in E \;\text{ and }\; \omega \notin F\}$$

Definition 1.6 *complement* The *complement* of E is denoted by E^c and is equal to $\Omega - E$.

Definition 1.7 (intersection of two sets) The *intersection* of two sets E and F is defined as

$$E \cap F = \{\omega : \ \omega \in E \ \text{ and } \ \omega \in F\}$$

Definition 1.8 (mutually exclsive) If $E \cap F = \emptyset$, E and F are said to be *disjoint* or *mutually exclusive*.

Definition 1.9 (union of two sets) The *union* of two sets E and F is defined as

$$E \cup F = \{\omega : \ \omega \in E \ \text{ or } \ \omega \in F\}$$

Definition 1.10 More generally if T is any set then

$$\cup_{t \in T} E_t = \{\omega : \omega \in E_t \ \text{ for some } \ t \in T\}$$

$$\cap_{t \in T} E_t = \{\omega : \omega \in E_t \ \text{ for all } \ t \in T\}$$

1.5.2 Properties of Set Operations

(1) $(E \cup F) \cup G = E \cup (F \cup G)$ and $(E \cap F) \cap G = E \cap (F \cap G)$

(2) $E \cup F = F \cup E$ and $E \cap F = F \cap E$

(3) $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$ and $E \cup (F \cap G) = (E \cup F) \cap (E \cup G)$

(4) $E \cup E^c = \Omega$ and $E \cap E^c = \emptyset$

(5) $(E \cup F)^c = E^c \cap F^c$

(6) $E - F = E \cap F^c$

(7) $(E^c)^c = E$

(8) $\Omega^c = \emptyset$ and $\emptyset^c = \Omega$

(9) $E \subset F \Rightarrow F^c \subset E^c$

(10) $(E_1 \times E_2) \cap (E_3 \times E_4) = (E_1 \cap E_3) \times (E_2 \cap E_4)$

(11) $(E \times F)^c = (E^c \times F^c) \cup (E \times F^c) \cup (E^c \times F)$

(12) $(\cup_{t \in T} E_t)^c = \cap_{t \in T} E_t$

(13) $(\cap_{t \in T} E_t)^c = \cup_{t \in T} E_t$

1.6 Counting and Combinatorics

1.6.1 Cartesian Products

Definition 1.11 (n-tuple) If $E_1, E_2, \ldots, E_n$ are sets an *n-tuple* is an element of the form

$$(a_1, a_2, \ldots, a_n) \text{ where } a_1 \in E_1, a_2 \in E_2, \ldots, a_n \in E_n$$

a_i is called the ith *coordinate* of the n-tuple.

Two n-tuples are said to be equal if each coordinate of one is equal to the corresponding coordinate of the other.

Note: An ordered pair with first coordinate a and second coordinate b is an element of the form (a, b). Also note that $(a, b) = (c, d)$ if and only if $a = c$ and $b = d$. n-tuples are natural generalizations of this concept. A formal (set) definition of an ordered pair is

$$(a, b) = \{\{a\}, \{a, b\}\}$$

.

Definition 1.12 (Cartesian product) If $E_1, E_2, \ldots, E_n$ are sets then their *Cartesian product* is defined as

$$E_1 \times E_2 \times \cdots \times E_n = \{(\omega_1, \omega_2, \ldots, \omega_n) : \omega_1 \in E_1, \omega_2 \in E_2, \ldots, \omega_n \in E_n\}$$

1.6.2 Cardinality

Definition 1.13 A set will be said to be:

(1) *finite* if its elements can be put into a one to one correspondence with the integers $1, 2, \ldots, n$ for some finite integer n.

(2) *denumerable* if its elements can be put into a one to one correspondence with the natural numbers N where $N = \{1, 2, 3, ...\}$

(3) *non-denumerable* otherwise.

1.6.3 Combinatorics: Elementary Counting Rules

Result 1.1 Elementary Counting Rules

(1) Given two sets A and B having m and n elements respectively there are $m \times n$ ordered pairs of the form (a_i, b_j) where $a_i \in A$ and $b_j \in B$

(2) Given r sets $A_1, A_2, \ldots, A_r$ containing $m_1, m_2, \ldots, m_r$ elements respectively, there are $m_1 \times m_2 \times \cdots \times m_r$ r-tuples of the form

$$(a_1, a_2, \ldots, a_r); \quad a_i \in A_i \;\; \text{for} \;\; i = 1, 2, \ldots, r$$

(3) A **permutation** of a set containing n elements is an arrangement of the elements of the set to form an n-tuple. There are n! permutations of a set containing n elements where $n!$ is defined as $n(n-1)\cdots 3 \cdot 2 \cdot 1$. By convention $0! = 1$. This convention is related to the Gamma function.

(4) Given a set containing n elements the number of subsets of size x is given by

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} = \frac{(n)_x}{x!}$$

where

$$(n)_x = (n-x+1)\cdots n = \prod_{i=0}^{x-1}(n-i)$$

This expression is read as n choose x and is called the number of **combinations** of n items taken x at a time.

(5) Given a population of size N there are

(i) N^n samples of size n with replacement.

(ii) $(N-n+1) \times \cdots \times N \equiv (N)_n \equiv N^{\underline{n}}$ samples of size n without replacement. The symbol $N^{\underline{n}}$ is called the falling factorial or Pochhammer's symbol.

1.6.4 Occupancy Numbers

Consider N urns and n balls. Assume the urns are distinguisable, numbered $1, 2, \ldots, N$

(i) The balls can be distinguishable (d) or not $(\overline{d})$

(ii) The urns can contain more than one ball, called an exclusion property (e) can contain at most one ball, ($\overline{e}$) can contain more than one ball.

The occupancy problem is to determine the number of ways in which the n balls can be placed (occupy) in the urns.

Case 1 : Distinguishable balls, no exclusion

$$O_{d,\overline{e}}(n, N) = N^n$$

Case 2 : Distinguishable balls, exclusion

$$O_{de}(n, N) = (N)_n$$

Case 3 : Non-distinguishable balls, no exclusion

$$O_{\overline{d},\overline{e}} = \binom{N+n-1}{n}$$

Case 4 : Non-distinguishable balls, exclusion

$$O_{\overline{d},e}(n < N) = \binom{N}{n}$$

All of the occupancy problems are easy except for Case 3. Here we argue that

N ways to place the first ball
$(N + 1)$ ways to place the second ball
....
$(N + n - 1)$ ways to place the nth ball

This assumes that the ordering within an urn is important. To remove this divide by $n!$ to get

$$O_{\overline{d},\overline{e}} = \frac{(N+n-1)_n}{n!} = \binom{N+n-1}{n}$$

1.7 Functions

Definition 1.14 If A and B are any two sets then any rule which assigns to each $x \in A$ a unique $y \in B$ is called a *function on A to B*.

(1) We denote a function on A to B by a letter, say f, and write

$$f : A \mapsto B$$

(2) The value of a function at a point $x \in A$ will be denoted by $f(x)$.

(3) The set A is called the *domain* of f and B is called the *codomain* of f.

(4) The *range* of f is the set $\mathcal{R}_f$ defined by

$$\mathcal{R}_f = \{y : y = f(x) \ \text{ for some } \ x \in A\}$$

(5) If $f : A \mapsto B$ and $g : B \mapsto C$ are two functions then the *composite* of g with f, $g \circ f$, is the function mapping A to C defined by $g \circ f(x) = g(f(x))$ for each $x \in A$

1.7.1 Graph of a Function

Definition 1.15 If f has domain A then the set of ordered pairs

$$\mathcal{G}(f) = \{(x, f(x) \ : \ x \in A\}$$

is called the *graph* of f.

1.7.2 Indicator Function

Definition 1.16 Indicator function The function I_E defined by

$$I_E(x) = \begin{cases} 1 & x \in E \\ 0 & x \notin E \end{cases}$$

is called the *indicator function* of the set E.

The following are some properties of indicator functions:

(1) $I_E \leq I_F$ if and only if $E \subset F$

(2) $E = F$ if and only if $I_E = I_F$

(3) $I_{E\cap F} = 0$ if and only if $E \cap F = \emptyset$

(4) $I_\emptyset = 0$, $I_\Omega = 1$

(5) $I_{E^c} = 1 - I_E$

(6) $I_{\cap_{i=1}^n E_i} = \prod_{i=1}^n I_{E_i}$

(7) $I_{\cup_{i=1}^n E_i} = \sum_{i=1}^n I_{E_i}$ if $E_i \cap E_j = \emptyset$ for $i \neq j$

2 Sequences and Series

Definition 2.1 A sequence of real numbers is a function from the set of natural numbers $N = \{1, 2, 3, \ldots\}$ to the set of real numbers. We denote a sequence of real numbers by $\{y_n : n \in N\}$ or simply by $\{y_n\}$. Similarly a sequence of complex numbers is denoted by $\{z_n : n \in N\}$ or by $\{z_n\}$

2.1 Limits of Sequences

A sequence is said to have a limit c as n tends to infinity if the terms of the sequence are "close" to c for large n. A more precise definition is the following:

Definition 2.2 A sequence of real numbers $\{y_n\}$ is said to converge to the limit c as n tends to infinity if for every $\epsilon > 0$ there exists an integer N_ϵ such that $n \geq N_\epsilon$ implies

$$| y_n - c | < \epsilon$$

We write $\lim_{n\to\infty} y_n = c$ or $y_n \to c$ when $\{y_n\}$ converges to c

Definition 2.3 A sequence of real numbers is said to be *monotone increasing* if $x_n \leq x_{n+1}$ for all n and *monotone decreasing* if $x_n \geq x_{n+1}$ for all n. In either case the sequence is said to be *monotone*.

Definition 2.4 A sequence of real numbers is bounded above (below) if there exists a real number B (b) such that $y_n < B$ for all n ($y_n > b$ for all n. The sequence is bounded if it is bounded above and below.

Results:

- A monotone increasing sequence (decreasing) sequence which is bounded above (below) has a limit.
- If $\lim_{n\to\infty} x_n = c$ and $\lim_{n\to\infty} y_n = d$ then

  $\lim(x_n + y_n) = c + d \quad \lim(x_n y_n) = cd$ if $c \neq 0$ and $y_n \neq 0$ for all n then $\lim 1/y_n) = 1/c$

- (Squeeze Theorem) If $x_n \leq y_n \leq z_n$ for all n and $\lim x_n = c$, $\lim z_n = c$ then $\lim y_n = c$

2.2 Some Particular Sequences

(1) For any real number a let $x_n = a^n$ for n an integer, then

  (a) x^n converges to 0 if $\mid a \mid < 1$

  (b) x_n converges to 1 if $a = 1$

  (c) x_n is bounded but does not converge if $a = -1$

  (d) x_n is not bounded and hence does not converge if $\mid a \mid > -1$

(2) (*geometric series*) For any real number a let $x_n = 1 + a + a^2 + \cdots + a^n$ for n an integer. Then x_n converges to $1/(1-a)$ if $\mid a \mid < 1$ and does not converge otherwise .

(3) For any real number a let $x_n = 1 + a + (a^2/2!) + \cdots + (a^n/n!)$ for n an integer, then x_n converges to e^a.

(4) For any real number a let $x_n = (1 + \frac{a}{n})^n$ for n an integer, then x_n converges to e^a.

2.3 o, O Notation for Sequences

Definition 2.5 Write

$$s_n \prec t_n \iff \lim_{n\to\infty} \frac{s_n}{t_n} = 0$$

We say that s_n is of smaller order than t_n. We also write

$$s_n = o(t_n) \text{ if } s_n \prec t_n$$

and say that s_n is "little o" of t_n.

Result 2.1 $\prec$ is transitive i.e.

$$\text{If } s_n \prec t_n \text{ and } t_n \prec u_n \text{ then } s_n \prec u_n$$

or

$$s_n = o(t_n) \text{ and } t_n = o(u_n) \implies s_n = o(u_n)$$

Result 2.2

$$1 \prec \ln[\ln(n)] \prec \ln(n) \prec n^{\epsilon} \prec n^{c} \prec n^{\ln(n)} \prec c^{n} \prec n^{n} \prec c^{c^{n}}$$

where $0 < \epsilon < 1 < c$

A useful notation for sequences is provided by the "little o"- "big O" notation.

Definition: If $\{a_n\}$ and $\{b_n\}$ are two sequences of real numbers we write

- $a_n = O(b_n)$ read as "a_n is big O of b_n" if there exists a positive number K and an integer N_K such that $n \geq N_K$ implies

$$| a_n | \leq K | b_n |$$

- $a_n = o(b_n)$ read as "a_n is little o of b_n" if for any $\epsilon > 0$ there exists an integer N_ϵ such that $n \geq N_\epsilon$ implies

$$| a_n | \leq \epsilon | b_n |$$

- We use o and O to compare the *magnitude* or *order* of two sequences.

2.4 Properties of o, O notation:

(1) A finite number of initial terms does not matter

(2) If c is a constant the statements

$$a_n = O(b_n) \text{ and } a_n = O(cb_n)$$
$$a_n = o(b_n) \text{ and } a_n = o(cb_n)$$

are equivalent. In particular the sign of an o or O term can be ignored.

(3) $a_n = o(1)$ means that $\lim_{n\to\infty} a_n = 0$

(4) $a_n = O(1)$ means that a_n is bounded.

(5) Multiplication of o and O terms obey the following rules

(a) $O(a_n)O(b_n) = O(a_n b_n)$

(b) $o(a_n)o(b_n) = o(a_n b_n)$

(c) $o(a_n)O(b_n) = o(a_n b_n)$

(6) When o and O terms are added, the order of magnitude of the sum is equal to the largest order of magnitude of the individual terms provided that the number of terms added does not depend on n. Thus

$$\begin{aligned} o(1) + o(n^{-1}) &= o(1) \\ o(n^{-1}) + o(n^{-2}) &= o(n^{-1}) \\ O(n^{-1}) + O(n^{-2}) &= O(n^{-1}) \end{aligned}$$

(7) The sequence $1/n + 1/n + \cdots + 1/n = a_n = 1$ shows that the order of magnitude of a sum is not provided by the largest order of magnitude of the individual terms (which is $1/n$) if the number of terms depends on n.

2.5 Infinite Series

Definition 2.6 Let $\{a_n\}$ be a sequence of real numbers. Then

$$\sum_{n=1}^{\infty} a_n$$

is called an infinite series.

Definition 2.7 $S_n = a_1 + a_2 + \cdots + a_n$ is called the nth partial sum of the infinite series.

(1) If $\lim_{n\to\infty} S_n = S$ the infinite series $\sum_{n=1}^{\infty} a_n$ is said to *converge* with sum S. Otherwise $\sum_{n=1}^{\infty} a_n$ is said to *diverge*.

(2) The sum, S, of an infinite series is unique.

(3) A series of (real or complex) converges absolutely if $\sum_{n=0}^{\infty} |z_n|$ converges.

Results:

- $e^x = \exp\{x\} = \sum_{n=0}^{\infty} x^n/n!$ converges absolutely for all x

2.6 Decimal and Binary Expansions of Numbers

Definition 2.8 The *decimal* expansion of a real number x in $[0,1]$ is

$$x = \sum_{n=1}^{\infty} \frac{a_n}{10^n} \text{ where } n \in \{0,1,2,\ldots,9\} \text{ for each } n$$

We usually write

$$x = .a_1 a_2 \ldots a_n \ldots$$

The decimal expansion is unique except for numbers of the form $k/10^n$ which have two expansions e.g.

$$.5 = .5000\ldots \quad \text{and} \quad .5 = .4999\ldots$$

Definition 2.9 The *binary* expansion of a real number x in the interval $[0,1]$ is

$$x = \sum_{n=1}^{\infty} \frac{a_n}{2^n} \text{ where } a_n \in \{0,1\} \text{ for each } n$$

Binary expansions are also essentially unique. They play an important role in probability theory as a model for an infinite sequence of coin tosses with the probability of a head $= 1/2$

3 Real Valued Functions, Limits and Continuity

Definition 3.1 A function whose domain is a set of real numbers (usually an interval) and whose range is also a set of real numbers will be called a real valued function.

3.1 Limits of Real Valued Functions

3.1.1 Limit of a Real Valued Function at a Point

Definition 3.2 The limit $L_f(x_0)$, written as

$$\lim_{\Delta \to 0} f(x_0 + \Delta)$$

of a function f at x_0, if it exists, is that number $L_f(x_0)$ such that $|f(x_0) - L_f(x_0)|$ is "small" whenever $|x - x_0|$ is small.

A more precise definition of limit is as follows:

Definition 3.3 If S is a set of real numbers a point c in S is said to be *adherent* to S if for every $\delta > 0$, there is an $x \in S$ satisfying $\mid x - c \mid < \delta$.

Definition 3.4 Let S be a set of real numbers, x_0 be adherent to S and let f map S into R. We say that $f(x)$ converges to ℓ as x tends to x_0 if for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$x \in S \;\text{ and }\; |x - x_0| < \delta \Rightarrow |f(x) - \ell| < \epsilon$$

We say that $f(x)$ tends to ℓ as x tends to x_0 or ℓ is the limit of f at x_0 or f has limit ℓ at c and write

$$\lim_{x \to x_0} f(x) = \ell$$

Theorem 3.1 f has limit ℓ at x_0 if and only if $\lim_{n\to\infty} f(x_n) = \ell$ for every sequence $\{x_n\}$ such that

$$\lim_{n \to \infty} x_n = x_0$$

3.1.2 Examples of Limits

In these examples the set S is the set of all real numbers and x_0 is any real number

(1) $f(x) = c = \text{constant}$	$L_f(x_0) = c$
(2) $f(x) = a + bx$, a, b constant	$L_f(x_0) = a + bx_0$
(3) $f(x) = e^x$	$L_f(x_0) = e^{x_0}$
(4) $f(x) = cos(x)$	$L_f(x_0) = cos(x_0)$
(5) $f(x) = sin(x)$	$L_f(x) = sin(x_0)$
(6) $f(x) = c\,h(x)$, c constant	$L_f(x_0) = cL_h(x_0)$
(7) $f(x) = g(x) \pm h(x)$	$L_f(x_0) = L_g(x_0) \pm L_h(x_0)$
(8) $f(x) = g(x)h(x)$	$L_f(x_0) = L_g(x_0)L_h(x_0)$
(9) $f(x) = g(x)/h(x)$	$L_f(x_0) = L_g(x_0)/L_h(x_0)$ provided $h(x) \neq 0$
(10) $f(x) = g \circ h(x)$	$L_f(x_0) = L_g(L_h(x_0))$

3.2 o, O Notation for Real Valued Functions

The "little o", "big O" notation can be used for real valued functions as the definition below shows.

Definition:

- $a(x) = O(b(x))$ as $x \to c$ if for any sequence $\{x_n\}$ such that $\lim x_n = x_0$ we have $a(x_n) = O(b(x_n))$.
- $a(x) = o(b(x))$ as $x \to x_0$ if for any sequence $\{x_n\}$ such that $\lim x_n = x_0$ we have $a(x_n) = o(b(x_n))$.

3.3 Continuous Real Valued Functions

3.3.1 Definitions

Definition 3.5 The function f is continuous at x_0 if $L_f(x_0) = f(x_0)$ A more precise definition follows.

Definition 3.6 Let S be a set of real numbers, $c \in S$ and let $f : S \mapsto R$. f is said to be continuous at x_0 if

$$\lim_{x \to x_0} f(x) = f(x_0)$$

i.e. if f is continuous at c if for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$|x - x_0| < \delta \;\Rightarrow\; |f(x) - f(x_0)| < \epsilon$$

Definition 3.7 A function is said to be continuous on S if it is continuous at every point in S.

3.4 Exponentials and Logarithms

Definition 3.8 An *exponential* is a function of the form $y = a^x$ where a is called the *base* and x is called the *exponent*.

If a is a positive constant the following defines the exponential function for any positive constant a

(i) If x is a positive integer n then

$$\underbrace{a \cdot a \cdots a}_{n\text{times}} = a^n$$

(ii) If $x = 0$ then $a^0 = 1$

(iii) If $x = -n$ where n is a positive integer then $a^{-n} = 1/a^n$

(iv) If x is rational i.e. $x = p/q$ where p and $q > 0$ are integers then $a^x = a^p/a^q$

(iv) a^x for any real number is then defined as $a^x = \lim_{r_n \to x} a^r$ where $r_n \to x$ is a sequence of rational numbers which converge to x

Properties of exponentials:

$$\begin{array}{cc} a^{x+y} = a^x a^y & a^{x-y} = \frac{a^x}{a^y} \\ (a^x)^y = a^{xy} & (ab)^x = a^x b^x \end{array}$$

Definition 3.9 If $a > 0$ and $a \neq 1$ the *logarithm to the base* a is defined by the equation

$$\log_a(x) = y \quad \longleftrightarrow \quad a^y = x$$

3.5 Examples and Properties of Continuous Functions

(1) $f(x) = c =$ constant
(2) $f(x) = a + bx$, a, b constant
(3) $f(x) = e^x$
(4) $f(x) = \log(x)$
(5) $f(x) = cos(x)$
(6) $f(x) = sin(x)$
(7) $f(x) = c\,h(x)$, c constant , $h(x)$ continuous
(8) $f(x) = g(x) \pm h(x)$, $g(x), h(x)$ continuous
(9) $f(x) = g(x)h(x)$, $g(x), h(x)$ continuous
(10) $f(x) = g(x)/h(x)$, $g(x), h(x)$ continuous provided $h(x) \neq 0$
(11) $f(x) = g \circ h(x)$, $g(x), h(x)$ continuous

4 Real Valued Functions - Derivatives

4.1 Loose Definition

Definition 4.1 The derivative of f at x_0, denoted by $f^{'}(x)$ or $Df(x_0)$, if it exists, is defined as

$$\lim_{\Delta \to 0} \frac{f(x_0 + \Delta) - f(x_0)}{\Delta}$$

4.2 Examples and Properties of Derivatives:

(1) $f(x) = c = \text{constant}$	$f^{'}(x) = 0$
(2) $f(x) = a + bx$, a, b constant	$f^{'}(x) = b$
(3) $f(x) = a + bx + cx^2$, a, b, c constant	$f^{'}(x) = b + 2cx$
(4) $f(x) = x^n$ $n = 1, 2, \ldots$	$f^{'}(x) = nx^{n-1}$
(5) $f(x) = x^{\alpha}$	$f^{'}(x) = \alpha x^{\alpha-1}$
(6) $f(x) = e^x$	$f^{'}(x) = e^x$
(7) $f(x) = \log(x)$	$f^{'}(x) = \frac{1}{x}$
(8) $f(x) = cos(x)$	$f^{'}(x) = -sin(x)$
(9) $f(x) = sin(x)$	$f^{'}(x) = cos(x)$
(10) $f(x) = g(x) \pm h(x)$	$L^{'}_{f}(x) = g^{'}(x) \pm h^{'}(x)$
(11) $f(x) = g(x)h(x)$	$f^{'}(x) = g^{'}(x)h(x) + g(x)h^{'}(x)$
(12) $f(x) = g(x)/h(x)$	$f^{'}(x) = \frac{g^{'}(x)}{h(x)} - \frac{g(x)h^{'}(x)}{[h(x)]^2}$ provided $h(x) \neq 0$
(13) $f(x) = g \circ h(x)$	$f^{'}(x) = g^{'}(h(x))h^{'}(x)$

4.3 Precise Definition of Derivative

Definition 4.2 The *interior* of an interval I of real numbers, int (I), is defined as

$$\text{int}\,(I) = \{x : \text{ there exists a } \delta > 0 \text{ such that } \{y : |y - x| < \delta\} \subset I\}$$

Definition 4.3 Let I be an interval of real numbers, $f : I \mapsto R$, $x_0 \in int(I)$, and define $g : I - \{x_0\} \mapsto R$ by

$$g(x) = \frac{f(x) - f(x_0)}{x - x_0}$$

(Note that x_0 is adherent to I) Then f has a *derivative*, $f'(x_0)$, at x_0 if

$$\lim_{x \to x_0} g(x) = f'(x_0)$$

Thus we have

Definition 4.4 f has a derivative $f'(x_0)$ at x_0 if for every $\epsilon > 0$ there exists a $\delta > 0$ such that such that $0 < |x - x_0| < \delta$ implies

$$\left| \frac{f(x) - f(x_0)}{x - x_0} - f'(x_0) \right| < \epsilon$$

We denote the derivative of f at x_0 by

$$f'(x_0) \text{ or } Df(x_0) \text{ or } f^{(1)}(x_0) \text{ or } \frac{df(x_0)}{dx}$$

and say that f is *differentiable* at x_0.

If $f : I \mapsto R$ and $A \subset I$ we say that f is differentiable on A if f has a derivative at every point c in A. The function $h : A \mapsto R$ defined by $h(x) = f'(x)$ will be called the derivative of f on A and will be denoted by

$$f', Df \text{ or } \text{or} f^{(1)} \text{ or } \frac{df}{dx}$$

4.4 Approximation Form of Derivative

Using "little o" notation the definition of a derivative may be rewritten as

$$\frac{f(x) - f(x_0)}{x - x_0} = f'(x_0) + o(1) \text{ as } x \to x + 0$$

Equivalently we may write

- $f(x) - f(x_0) = f'(x_0)(x - x_0) + (x - x_0)o(1)$
- $f(x) = f(x_0) + f'(x_0)(x - x_0) + o(|x - x_0|)$

That is, f can be approximated at x by a linear expression of the form

$$f(x) = a + bx \text{ where } a = f(x_0) - f'(x_0)x_0 \text{ and } b = f'(x_0)$$

4.5 Derivatives of Higher Order

Let I be an interval and let $f : I \mapsto R$. If f is differentiable on I denote its derivative by Df or f'. If Df is differentiable on I denote its derivative by D^2f i.e. $D(Df) = D^2f$ or by $f^{(2)}$. The nth derivative of f on I if it exists will be denoted by D^nf or $f^{(n)}$.

4.6 Maxima and Minima

4.6.1 Loose Criteria

Let $I \subset R$, $f : I \mapsto R$, and $c \in I$. Then f

- Has a minimum at c if $f^{(1)}(c) = 0$ and $f^{(2)}(c) > 0$
- Has a maximum at c if $f^{(1)}(c) = 0$ and $f^{(2)}(c) < 0$

4.6.2 Precise Criteria

1. Let $I \subset R$, $f : I \mapsto R$, and $c \in I$. Then

   (a) f has a *relative strict minimum* at c if there is an $\epsilon > 0$ such that

   $$x \in I \text{ and } |x - c| < \epsilon \Rightarrow f(c) \leq f(x)$$

   (b) f has a *relative strict maximum* at c if there is an $\epsilon > 0$ such that

   $$x \in I \text{ and } |x - c| < \epsilon \Rightarrow f(c) \geq f(x)$$

2. *First Derivative Test:* Let $I \subset R$ be an interval, $f : I \mapsto R$ suppose that f is differentiable on I and that $c \in \text{int}\,(I)$. Then

   (a) c is a point of relative strict minimum of f if there is an $\alpha > 0$ such that

   $$Df(x) < 0 \text{ for } x \in (c - \alpha, c) \text{ and } Df(x) > 0 \text{ for } x \in (c, c + \alpha)$$

   (b) c is a point of relative strict maximum of f if there is an $\alpha > 0$ such that

   $$Df(x) > 0 \text{ for } x \in (c - \alpha, c) \text{ and } Df(x) < 0 \text{ for } x \in (c, c + \alpha)$$

3. *Second Derivative Test:* Let $I \subset R$ be an interval, $f \in C^{(2)}(I)$ and let $c \in I$. Then

   (a) c is a point of relative strict minimum if $Df(c) = 0$ and $D^2f(c) > 0$.

   (b) c is a point of relative strict maximum if $Df(c) = 0$ and $D^2f(c) < 0$.

4.7 Taylor's Theorem and the Mean Value Theorem

4.7.1 Most Used Form

$$f(x) = f(x_0) + f^{(1)}(x_0)(x - x_0) + o(|x - x_0|)$$

4.7.2 Precise Form

Theorem 4.1 Mean Value Theorem: Let I be an interval and suppose that $[a, b] \subset I$. If $f \in C(I)$ is differentiable on $int(I)$ then there is a point $c \in (a, b)$ such that

$$f(b) - f(a) = f'(c)(b - a)$$

Theorem 4.2 Taylor's Theorem: Let f have r derivatives at $x = x_0$. Then

$$\begin{aligned} f(x) &= f(x_0) + f'(x_0)h + \cdots + \frac{f^{(r)}(x_0)(x - x_0)^r}{r!} + o(|x - x_0|^r) \\ \text{or } f(x) &= \sum_{i=0}^{r} \frac{f^{(i)}(x_0)(x - x_0)^i}{i!} + o(|x - x_0|^r) \end{aligned}$$

which we may also write as

$$\left| f(x) - \sum_{i=0}^{r} \frac{f^{(i)}(x_0)(x - x_0)^i}{i!} \right| \leq o(|x - x_0|^r)$$

4.8 Taylor's Series

Theorem 4.3 Taylor's Series Let f have derivatives of all orders in an interval $(x_0 - r, x_0 + r)$ and assume that these derivatives are bounded i.e.

$$|f^{(n)}(x)| \leq B \text{ for all } x \in (x_0 - r, x_0 + r)$$

Then for each $x \in (x_0-, x_0 + r)$ we have

$$f(x) = \sum_{i=0}^{\infty} \frac{f^{(i)}(x_0)(x - x_0)^i}{i!}$$

4.9 Power Series

Definition 4.5 A **power series** in $x - x_0$ is an infinite series of the form

$$f(x) = \sum_{i=0}^{\infty} a_i (x - x_0)^i$$

The numbers $a_0, a_1, a_2, \ldots$ are called its *coefficients*. If this seires converges for $x \in (x_0 - r, x_0 + r)$ the r is called the *radius of convergence* of the power series.

The most important example of a power series is

$$e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$$

which is valid for all real x.

The most important property of power series is that they uniquely determine the coefficients in the radius of convergence. That is, if you know the power series $f(x)$ then you know the coefficients and conversely.

4.10 L'Hospital's Rule

In obtaing limits of fractions the following result is often useful.

Theorem 4.4 L'Hospital's Rule If $\lim_{x \to c} f(x) = \lim_{x \to c} g(x) = 0$ then

$$\lim_{x \to c} \frac{f(x)}{g(x)} = \lim_{x \to c} \frac{f'(x)}{g'(x)}$$

if the limit on the right exists.

A similar conclusion holds if

$$\lim_{x \to c} f(x) = \lim_{x \to c} g(x) = \infty$$

5 Integration

5.1 Definitions

Definition 5.1 Let I be an interval of real numbers and let $f : I \subset R$. A *primitive* or *indefinite integral* of f on I is any function $F : I \mapsto R$ such that $DF(x) = f(x)$ for all $x \in I$.

Definition 5.2 If I is an interval of real numbers, f is continuous on I and F is a primitive of f on I then for every $a \in I$ and $b \in I$ the difference $F(b) - F(a)$ is called the *integral* of f from a to b and is denoted by

$$F(b) - F(a) = \int_a^b f(x)dx$$

Definition 5.3 If A is a set of real numbers and $f : A \mapsto R$ is non-negative for $x \in A$ we call the set

$$O_f = \{(x, y) : x \in A, 0 \leq f(x) \leq y\}$$

the *ordinate set* of f.

Theorem 5.1 Fundamental Theorem of the Calculus Let I be an interval of real numbers and let f be continuous on I. Then f has a primitive on I. Moreover, if $I = [a, b]$, $f(x) \geq 0$ for $x \in I$ and $A(O_f)$ denotes the area of O_f then

$$A(O_f) = \int_a^b f(x)dx$$

Note the folowing properties of integrals:

$$\begin{array}{rcl}
\int_a^b f(x)dx & = & \int_a^b f(t)dt \\
\int_a^a f(x)dx & = & 0 \\
\int_a^b f(x)dx & = & -\int_b^a f(x)dx \\
\int_a^b f(x)dx + \int_b^c f(x)dx & = & \int_a^c f(x)dx
\end{array}$$

5.2 Examples and Properties of Integrals

$$
\begin{array}{clrcl}
(1) & f(x) = c & \int f(x)dx & = & cx \\
(2) & f(x) = x^n & \int f(x)dx & = & \frac{x^{n+1}}{n+1} \;\; n \neq -1 \\
(3) & f(x) = x^\alpha & \int f(x)dx & = & \frac{x^{\alpha+1}}{\alpha+1} \;\; \alpha \neq -1 \\
(4) & f(x) = cg(x) & \int f(x)dx & = & c\int g(x)dx \\
(5) & f(x) = g(x) \pm h(x) & \int f(x)dx & = & \int g(x)dx \pm \int h(x)dx \\
(6) & f(x) = \frac{1}{x} & \int f(x)dx & = & \log(x) \;\; \text{base } e \\
(7) & f(x) = e^x & \int f(x)dx & = & e^x \\
(8) & f(x) = sin(x) & \int f(x)dx & = & -cos(x) \\
(9) & f(x) = cos(x) & \int f(x)dx & = & sin(x)
\end{array}
$$

5.3 Integration Methods

Theorem 5.2 Substitution Method Let ϕ be a real valued function on the closed bounded interval $[a, b]$ with a derivative ϕ' which is also continuous on $[a, b]$. Let $\psi(x)$ be the inverse of $\phi(x)$. If f is continuous on $\phi([a, b])$ then

$$\int_a^b f[\phi(x)]dx = \int_{\phi(a)}^{\phi(b)} f(u)\psi'(u)du$$

Theorem 5.3 Itegration by Parts Let I be na interval, f and g be continuous on I. If a and b are in I then

$$\int_a^b f(x)g'(x)dx = f(x)g(x)\Big|_a^b - \int_a^b f'(x)g(x)dx$$

We usually write $u = f(x)$ and $dv = g'(x)dx$. Then the formula for integration by parts becomes

$$\int udv = uv - \int vdu$$

5.4 Liebnitz's Rule

In many distribution problems in statistics we need to find the derivative of an integral.

Theorem 5.4 Liebnitz's Rule Let

$$H(\alpha) = \int_{h_1(\alpha}^{h_2(\alpha)} h(x;\alpha)dx$$

then

$$h^{'}(\alpha) = \int_{h_1(\alpha}^{h_2(\alpha)} \frac{\partial h(x;\alpha)}{\partial \alpha} dx + h_2^{'}(\alpha)h[h_2(\alpha),\alpha] - h_1^{'}(\alpha)h[h_1(\alpha),\alpha)]$$

5.5 Logarithms and Exponentials

5.6 Logarithmic function

Define the logarithmic function $\log : (0,\infty) \mapsto R$ (also called the natural log) by

$$\log(x) = \int_1^x \frac{1}{y} dy \ \text{ for } x \in R$$

Properties

- log is continuous and strictly increasing for $x \in (0,\infty)$
- log is differentiable on $(0,\infty)$, in fact
$$\frac{d\log(x)}{dx} = \frac{1}{x}$$
- $\log(xy) = \log(x) + \log(y)$
- $\log(x/y) = \log(x) - \log(y)$
- $\log(1) = 0;\ \log(e) = 1$
- $\lim_{x\to\infty} \log(x) = \infty;\ \lim_{x\to 0} \log(x) = -\infty$
- $$\log(1+x) = \sum_{r=1}^{\infty} \frac{(-1)^{r+1}x^r}{r} = x - \frac{x}{2} + \frac{x^3}{3} - \ldots \ \text{ for } x \in (-1,+1]$$

- $$\log\left(\frac{1}{1-x}\right) = \sum_{r=1}^{\infty} \frac{x^r}{r} = x + \frac{x}{2} + \frac{x^3}{3} + \dots \quad \text{for } x \in [-1, +1)$$

- $$x - \frac{x^2}{2} + \cdots - \frac{x^{2n}}{2n} < \log(1+x) < x - \frac{x^2}{2} + \frac{x^{2n+1}}{2n+1} \quad \text{for } x > 0$$

5.7 Exponential Function

Let $\exp : R \mapsto (0, \infty)$ be defined by

$$\log(\exp(x)) = x$$

Properties

- exp is continuous on R
- exp is differentiable on R, in fact

$$\frac{d\exp(x)}{dx} = x$$

- $\exp(0) = 1$, $\exp(x) > 0$ for $x \in R$
- $\exp(x+y) = \exp(x)\exp(y)$ for $x, y \in R$
- $\exp(-x) = [\exp(x)]^{-1}$ for $x \in R$
- $\exp(nx) = [\exp(x)]^n$ for $x \in R$
- $\lim_{x\to\infty} \exp(x) = +\infty$; $\lim_{x\to-\infty} \exp(x) = 0$
- $\exp(x) = \lim_{n\to\infty}[1 + \frac{1}{n}]^n$
- $\exp(x) = \sum_{r=1}^{\infty} x^r/r!$
- We usually write $\exp(x) = e^x$ and then

$$e^{\log(x)} = x \ ; \ \log(e^x) = x$$

6 n-Dimensional Calculus

6.1 n-Dimensional Euclidean Space

Definition 6.1 R^n will denote n-dimensional Euclidean space i.e. the set of all n-tuples of the form $(x_1, x_2, \ldots, x_n)$ where x_i is a real number for $i = 1, 2, \ldots, n$.

A point in R^n is called a vector and we write $\mathbf{x}$ to denote such a point. For a vector $\mathbf{x}$, x_1 is said to be the first coordinate, x_2 the second coordinate and so on.

Multiplication of a vector by a real number and addition of two vectors (having the same number of coordinates) is coordinate-wise i.e. $\lambda\mathbf{x}$ has ith coordinate λx_i and $\mathbf{x}+\mathbf{y}$ has ith coordinate $x_i + y_i$.

Definition 6.2 The **distance** between two vectors $\mathbf{x}$ and $\mathbf{y}$ is be

$$d(\mathbf{x}, \mathbf{y}) = [(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2]^{1/2}$$

Definition 6.3 The **norm** or **length** of a vector $\mathbf{z}$, $||\mathbf{z}||$ is

$$||\mathbf{z}||^2 = z_1^2 + z_2^2 + \cdots + z_n^2$$

Then $d(\mathbf{x}, \mathbf{y}) = ||\mathbf{x} - \mathbf{y}||$ and we note that

(1) $d(\mathbf{x}, \mathbf{y}) = 0$ if and only if $\mathbf{x} = \mathbf{y}$

(2) $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$

(3) $d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y})$ for any $\mathbf{z}$

Definition 6.4 More generally the **inner product** of two vectors $\mathbf{x}$ and $\mathbf{y}$ is

$$(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} x_i y_i$$

Thus $||\mathbf{x}||^2 = (\mathbf{x}, \mathbf{x})$.

Result 6.1 Inner products have the following properties:

(1) $(\mathbf{x}, \mathbf{x}) \geq 0$ with equality holding if and only if $\mathbf{x} = \mathbf{0}$.

(2) $(\mathbf{x}, \mathbf{y}) = (\mathbf{y}, \mathbf{x})$

(3) $(a\mathbf{x} + b\mathbf{y}, \mathbf{z}) = a(\mathbf{x}, \mathbf{z}) + b(\mathbf{y}, \mathbf{z})$

(4) $|(\mathbf{x}, \mathbf{y})| \leq ||\mathbf{x}||||\mathbf{y}||$ (Cauchy-Schwartz Inequality)

Definition 6.5 Two vectors are said to be **orthogonal** if $(\mathbf{x}, \mathbf{y}) = 0$.

The rationale behind this definition is that the angle θ between two vectors is defined as

$$\cos(\theta) = \frac{(\mathbf{x}, \mathbf{y})}{||\mathbf{x}||||\mathbf{y}||}$$

and the cosine of a right angle is 0.

6.2 Differentiation in R^n

Definition 6.6 In R^n the vectors $\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_n}$ where

$$\mathbf{e_i} = (\delta_{1i}, \delta_{2i}, \ldots, \delta_{ni})$$

and

$$\delta_{ji} = \begin{cases} 1 & j = i \\ 0 & j \neq i \end{cases}$$

is called the **natural** or **canonical** basis.

The reason for the term basis is that every $\mathbf{x} \in R^n$ can be represented as

$$\mathbf{x} = \sum_{i=1}^{n} x_i \mathbf{e_i}$$

6.2.1 Directional Derivatives

Definition 6.7 Any vector $\mathbf{h} \in R^n$ such that $||\mathbf{h}|| = 1$ is called a **direction** in R^n.

Thus $\mathbf{e_1}$ is the first coordinate direction, $\mathbf{e_2}$ the second coordinate direction and so on.

Definition 6.8 Given $\mathbf{x_0}$ and a direction $\mathbf{h}$ the **line through $\mathbf{x_0}$ with direction h** is the set

$$L(\mathbf{x_0}, \mathbf{h}) = \{\mathbf{x} : \mathbf{x} = \mathbf{x_0} + t\mathbf{h},\ t \in R\}$$

Definition 6.9 Let $S \subset R^n$, $\mathbf{x_0}$ be an interior point of S and let $f : S \mapsto R$. The directional derivative of f at $\mathbf{x_0}$ with direction $\mathbf{h}$ is

$$\lim_{t \to 0} \frac{[f(\mathbf{x_0} + t\mathbf{h}) - f(\mathbf{x_0})]}{t}$$

provided the limit exists.

6.2.2 Partial Derivatives

Let $S \subset R^n$, $\mathbf{x}$ be an interior point of S and let $f : S \mapsto R$

Definition 6.10 *The* **partial derivatives** *of f at $\mathbf{x_0}$ are the directional derivatives of f at $\mathbf{x_0}$ in the directions $\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_n}$ if they exist.*

The the ith partial derivative of f is

$$\lim_{t \to 0} \frac{f(x_1, x_2, \ldots, x_{i-1}, x_i + t, x_{i+1}, \ldots, x_n) - f(x_1, x_2, \ldots, x_n)}{t}$$

if the limit exists.

Notation: The usual notation for the ith partial derivative of f at $\mathbf{x_0}$ is

$$D_i f(\mathbf{x_0}) \text{ or } \partial f(\mathbf{x_0})/\partial x_i$$

Theorem 6.1 The Chain Rule If f is a function of $x_1, x_2, \cdots, x_n$ and each x_i is a function of t then the derivative of f with respect to t is given by

$$\frac{d\,f}{d\,t} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \frac{d\,x_i}{d\,t}$$

6.2.3 Higher Order Partial Derivatives

(1) The partial derivatives $D_1 f(\mathbf{x}), D_2 f(\mathbf{x}), \ldots, D_n f(\mathbf{x})$ may themselves possess partial derivatives. Such a partial derivative is called a partial derivative of order 2 and is denoted by $D_{ij} f(\mathbf{x})$ or by $\partial^2 f(\mathbf{x})/\partial x_i x_j$

(2) In general $D_{i_1,i_2,\ldots,i_n} f(\mathbf{x})$ will denote the partial derivative of order $i_1 + i_2 + \cdots + i_n$

(3) If all partial derivatives of a given order, say m, exist for each $\mathbf{x} \in S$ and if they are continuous on S we say that f belongs to the class $C^{(m)}(S)$.

(4) If $f \in C^{(2)}(S)$ and if $D_{ij} f$ and $D_{ji} f$ are continuous for $i \neq j$ then $D_{ij} f = D_{ji} f$.

6.3 Taylor's Theorem in R^n

6.3.1 Most Used Form

$$f(x,y) = f(x_0,y_0) + \frac{\partial f(x_0,y_0)}{\partial x}(x - x_0) + \frac{\partial f(x_0,y_0)}{\partial y}(y - y_0) + o(\sqrt{(x-x_0)^2 + (y-y_0)^2})$$

6.3.2 General Formulation

If f is a function of $x_1, x_2, \ldots, x_n$ with continuous partial derivatives then

$$f(x_1, x_2, \ldots, x_n) = f(a_1, a_2, \ldots, a_n) + \sum_{i=1}^{n} \frac{\partial f(\mathbf{a})}{\partial x_i}(x_i - a_i) + o(||\mathbf{x} - \mathbf{a}||)$$

More generally we have: Let $\mathbf{k}$ be an n-tuple of non-negative integers $k_1, k_2, \ldots, k_n$ and define

(1) $|\mathbf{k}| = \sum_{i=1}^{n} k_i$

(2) $\mathbf{h}^{\mathbf{k}} = h_1^{k_1} h_2^{k_2} \cdots h_n^{k_n}$ for $\mathbf{h} \in R^n$.

(3) $\mathbf{k}! = k_1! k_2! \cdots k_n!$ and $D^{\mathbf{k}} = D^{k_1} D^{k_2} \cdots D^{k_n}$.

Then if f has a continuous $(r+1)$st differential on an interval I containing the line segment LS joining $\mathbf{a}$ and $\mathbf{a}+\mathbf{h}$ there is a point $\boldsymbol{\theta} \in LS$ such that

$$
\begin{aligned}
f(\mathbf{a}+\mathbf{h}) &= \sum_{i=0}^{r-1} \frac{1}{i!} \left[\sum_{\mathbf{k}:|\mathbf{k}|=i} \binom{i}{\mathbf{k}} D^{\mathbf{k}} f(\mathbf{a}) \mathbf{h}^{\mathbf{k}} \right] + \sum_{\mathbf{k}:|\mathbf{k}|=r} \binom{r}{\mathbf{k}} \frac{D^{\mathbf{k}} f(\boldsymbol{\theta}) \mathbf{h}^{\mathbf{k}}}{r!} \\
&= \sum_{i=0}^{r-1} \frac{1}{i!} \sum_{\mathbf{k}:|\mathbf{k}|=i} \binom{i}{\mathbf{k}} D^{\mathbf{k}} f(\mathbf{a}) \mathbf{h}^{\mathbf{k}} + o(||\mathbf{h}||^r)
\end{aligned}
$$

where

$$\binom{i}{\mathbf{k}} = \frac{i!}{\mathbf{k}!(i-|\mathbf{k}|)!}$$

for $i = 0, 1, \ldots, r$ and $D^0 f(\mathbf{a}) = f(\mathbf{a})$

6.4 Maxima and Minima

6.4.1 Unconstrained Maxima and Minima

Definition 6.11 The **gradient** of f at $\mathbf{c}$ is

$$\nabla f(\mathbf{c}) = [\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \ldots, \frac{\partial f(\mathbf{x})}{\partial x_n}]|_{\mathbf{x}=\mathbf{c}}$$

Definition 6.12 The **Hessian** of f at $\mathbf{c}$ is

$$H_f(\mathbf{c}) = det[D_{ij}(f)\mathbf{c})]$$

where

$$D_{ij} f(\mathbf{c}) = \left. \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} \right|_{\mathbf{x}=\mathbf{c}}$$

and det is the determinant function.

Result 6.2 Let $U \subset R^n$ be an "open interval" of R^n and suppose that $f : U \mapsto R$ is differentiable on U and has a local minimum or maximum at $\mathbf{c}$. Then $\mathbf{c}$ is said to be a *critical point* of f and $\nabla f(\mathbf{c}) = 0$.

Result 6.3 Let $U \subset R^n$ be an "open interval" of R^n, suppose that $f : U \mapsto R$ is in $C^{(3)}(U)$ and that $\mathbf{c} \in U$ is a critical point of f. Then f has

(1) A local minimum at $\mathbf{c}$ if the Hessian, $H_f(\mathbf{c}) > 0$.

(2) A local maximum at $\mathbf{c}$ if the Hessian, $H_f(\mathbf{c}) < 0$.

(3) Neither if the Hessian is indefinite.

6.4.2 Lagrange Multipliers

Result 6.4 Let $g : R^n \mapsto R$ be continuously differentiable and let M be the set of points in R such that $g(x) = 0$ and $\nabla g(\mathbf{x}) \neq 0$. If the differentiable function $f : R^n \mapsto R$ attains a local minimum or maximum at $\mathbf{c} \in M$ then for some λ (called a Lagrange Multiplier)

$$\nabla f(\mathbf{c}) = \lambda \nabla g(\mathbf{c})$$

Thus to minimize or maximize a function subject to the constraint that $g(\mathbf{x}) = 0$ one solves the equations

$$g(\mathbf{c}) = 0 \;\text{ and }\; \nabla f(\mathbf{c}) = \lambda g(\mathbf{c})$$

for $c_1, c_2, \ldots, c_n$ and λ. The values so obtained give the critical point $c_1, c_2, \ldots, c_n$ such that $g(c_1, c_2, \ldots, c_n) = 0$.

For several restrictions one solves the system of equations:

$$\begin{aligned} \nabla f(\mathbf{c}) &= \sum_{i=1}^{m} \lambda_i g_i(\mathbf{c}) \\ g_i(\mathbf{c}) &= 0 \;\text{ for }\; i = 1, 2, \ldots, m \end{aligned}$$

for $c_1, c_2, \ldots, c_n, \lambda_1, \lambda_2, \ldots, \lambda_m$. Then $c_1, c_2, \ldots, c_n$ gives the critical point which satisfies the restrictions defined by $g_1, g_2, \ldots, g_n$.

6.5 Integration in R^n

Just as the area under the curve of a positive continuous function is given by

$$\int f(x)dx$$

the volume under the surface of a positive function of two variables is given by

$$\int\int f(x,y)dxdy$$

This process can be continued to obtain

$$\int\int\cdots\int f(x,y,\ldots,z)dxdy\cdots dz$$

the n dimensional integral.

Fortunately integration in such situations can be handled one variable at a time.

Result 6.5 Let f be defined and bounded on the rectangle $R=[a,b]\times[c,d]$. For each fixed y in $[c,d]$ assume that the one-dimensional integral

$$A(y)=\int_a^b f(x,y)dx$$

exists. Then the integral

$$\int_c^d A(y)dy$$

exists and is equal to the double integral of $f(x,y)$ over R i.e.

$$\int_c^d \left\{\int_a^b f(x,y)dx\right\} dy = \int_c^d \int_a^b f(x,y)dx\, dy$$

This is called an **iterated integral**. It allows multidimensional integrals to be evaluated by repeated single integration.

7 Matrices

7.1 Basics

7.1.1 Introduction

Definition 7.1 A matrix $\mathbf{L}$ is an $n\times p$ array of scalars (numbers). The number of rows is n and the number of columns is p. $n\times p$ is called the order of the matrix.

noindent **examples:**

$$\mathbf{L}_1 = \begin{bmatrix} 3 & 1 & 0 \\ 2 & 0 & 4 \\ 5 & 1 & 3 \end{bmatrix} \quad \mathbf{L} = \begin{bmatrix} 1 & 0 \\ 2 & 3 \\ 2 & 1 \end{bmatrix}$$

are 3×3 and 3×2 matrices.

Notation:

- The numbers making up the matrix are called the elements of the matrix.
- We write $\mathbf{L} = \{\ell_{ij}\}$ and call ℓ_{ij} the i, j element of $\mathbf{L}$ (Thus ℓ_{ij} is the element in the ith row and jth column of the matrix $\mathbf{L}$).
- Two matrices are equal, $\mathbf{A} = \mathbf{B}$, if and only if $a_{ij} = b_{ij}$ for all i and j.
- The zero or null matrix is the matrix each of whose elements is 0 and is denoted as $\mathbf{0}$.
- If $n = p$ the matrix $\mathbf{A}$ is said to be a square matrix.
- The **identity** matrix $\mathbf{I}$ is defined by

$$\mathbf{I} = (\delta_{ij}) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

  Thus the jth column of the identity matrix consists of a one in the jth row and zeros elsewhere.

- A $1 \times p$ matrix is called a **row vector** and is written as

$$\boldsymbol{\ell}_i^\top = (\ell_{i1}, \ell_{i2}, \ldots, \ell_{ip})$$

- Similarly an $n \times 1$ matrix is called a **column vector** and is written as

$$\boldsymbol{\ell}_j = \begin{bmatrix} \ell_{1j} \\ \ell_{2j} \\ \vdots \\ \ell_{nj} \end{bmatrix}$$

- With this notation we can write the matrix $\mathbf{L}$ in terms of its rows as

$$\mathbf{L} = \begin{bmatrix} \boldsymbol{\ell}_1^\top \\ \boldsymbol{\ell}_2^\top \\ \vdots \\ \boldsymbol{\ell}_n^T \end{bmatrix}$$

- Alternatively if $\boldsymbol{\ell}_j$ denotes the jth column of $\mathbf{L}$ we may write the matrix $\mathbf{L}$ in terms of its columns as

$$\mathbf{L} = [\boldsymbol{\ell}_1,\ \boldsymbol{\ell}_2,\ \cdots\ ,\boldsymbol{\ell}_p]$$

7.1.2 Sums and Products of Matrices

Definition 7.2 The sum of two matrices $\mathbf{A}$ and $\mathbf{B}$ is defined as

$$\mathbf{A} + \mathbf{B} = \{a_{ij} + b_{ij}\}$$

Note that $\mathbf{A}$ and $\mathbf{B}$ both must be of the same order for the sum to be defined.

Definition 7.3 The multiplication of a matrix by a scalar is defined by the equation

$$\lambda\mathbf{A} = \{\lambda\, a_{ij}\}$$

Matrix and addition and scalar multiplication of a matrix have the following properties:

- $\mathbf{A} + \mathbf{O} = \mathbf{O} + \mathbf{A} = \mathbf{A}$
- $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$
- $\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$
- $\mathbf{A} + (-\mathbf{A}) = \mathbf{O}$

Definition 7.4 The product $\mathbf{C}$ of two matrices $\mathbf{A}$ and $\mathbf{B}$ is defined by the equation

$$\mathbf{AB} = \mathbf{C} = \{c_{ij}\} = \{\sum_k a_{ik}b_{kj}\}$$

- The matrix of the product can be found by taking the ith row of $\mathbf{A}$ times the jth column of $\mathbf{B}$ element by element and summing.
- Note that the product is defined only if the number of columns of $\mathbf{A}$ is equal to the number of rows of $\mathbf{B}$.
- We say that $\mathbf{A}$ premultiplies $\mathbf{B}$ or that $\mathbf{B}$ post multiplies $\mathbf{A}$ in the product $\mathbf{AB}$. Matrix multiplication is not commutative. That is, $\mathbf{AB}$ does not equal $\mathbf{BA}$. In fact $\mathbf{BA}$ may not be defined.

Provided the indicated products are defined matrix multiplication has the following properties:

- $\mathbf{AO} = \mathbf{O}$ and $\mathbf{OA} = \mathbf{O}$
- $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC}$
- $(\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC}+\mathbf{BC}$
- $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$
- The identity matrix has the property that if A is a square matrix then
$$\mathbf{AI} = \mathbf{IA} = \mathbf{A}$$

7.1.3 Transposes

Definition 7.5 The **transpose** of $\mathbf{A}$, $\mathbf{A}^\top$ is
$$\mathbf{A}^\top = \{a_{ji}\}$$

Thus if $\mathbf{A}$ is n by p the transpose $\mathbf{A}^\top$ is p by n with i,j element equal to the j,i element of $\mathbf{A}$. To find the transpose of a matrix simply write the elements of each row as the columns of a new matrix. The resulting matrix is the transpose.

The following are some properties of the transpose operation:
$$\begin{array}{ccc} (\mathbf{AB})^\top & = & \mathbf{B}^T\mathbf{A}^T \\ \mathbf{I}^\top & = & \mathbf{I} \\ (\mathbf{A}+\mathbf{B})^\top & = & \mathbf{A}^T+\mathbf{B}^T \\ (\alpha\mathbf{A})^\top & = & \alpha\mathbf{A}^T \end{array}$$

Definition 7.6 A **column vector** is an n by one matrix and a **row vector** is a one by n matrix. If $\mathbf{x}$ is a column vector then $\mathbf{x}^\top$ is the row vector with the same elements.

If $\mathbf{a}^\top$ is a row vector of length n and $\mathbf{b}$ is a column vector of length n then the product $\mathbf{a}^\top\mathbf{b}$ is a 1 by 1 matrix i.e. a number and is given by $\sum_{i=1}^n a_i b_i$. This result gives another way of representing the product of $\mathbf{A}^\top\mathbf{B}$. If $\mathbf{a}_i^T$ is the ith row of $\mathbf{A}$ and $\mathbf{B}$ is the jth row of $\mathbf{B}$ then the $i-j$ element of the product $\mathbf{AB}$ is $\mathbf{a}_i^\top\mathbf{b}_j$.

7.1.4 Some Special Matrices

Definition 7.7 A square matrix $\mathbf{A}$ is said to be:

- **symmetric** if $\mathbf{A}^\top = \mathbf{A}$
- **diagonal** if $a_{ij} = 0 \text{ for } i \neq j$
- **upper right triangular** if $a_{ij} = 0 \text{ for } i > j$
- **lower left triangular** if $a_{ij} = 0 \text{ for } i < j$
- **idempotent** if $\mathbf{A}^2 = \mathbf{A}$
- **orthogonal** if $\mathbf{A}^\top\mathbf{A} = \mathbf{A}\mathbf{A}^T = \mathbf{I}$

examples:

(1)

$$B = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \text{ is symmetric}$$

(2)

$$B = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \text{ is diagonal}$$

(3)

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 2 & 0 \\ 4 & 1 & 2 \end{bmatrix} \text{ is lower triangular}$$

(4)

$$B = \begin{bmatrix} 5 & 4 & 1 \\ 0 & 2 & 3 \\ 0 & 0 & 6 \end{bmatrix} \text{ is upper triangular}$$

(5)

$$B = \begin{bmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{bmatrix} \text{ is idempotent}$$

(6)

$$B = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \text{ is orthogonal}$$

7.1.5 Partitioned Matrices

If $\mathbf{A}$ is $n \times p$ and is written as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

then $\mathbf{A}$ is said to be a partitioned matrix.

Result:

$$\mathbf{A}^\top = \begin{bmatrix} \mathbf{A}_{11}^\top & \mathbf{A}_{21}^T \\ \mathbf{A}_{12}^T & \mathbf{A}_{22}^T \end{bmatrix}$$

Result: If B is also n by p and is similarly partitioned then

$$A + B = \begin{bmatrix} A_{11} + B_{11} & A_{12} + B_{12} \\ A_{21} + B_{21} & A_{22} + B_{22} \end{bmatrix}$$

Result: If $\mathbf{A}$ $(n \times p)$ and $\mathbf{B}$ $p \times m$ are written as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}$$

then the product $\mathbf{AB}$ satisfies

$$\mathbf{AB} = \begin{bmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{bmatrix}$$

if the multiplications are defined.

7.1.6 Inverses

Definition: A square matrix $\mathbf{A}$ is said to be **invertible** or have an inverse if

$$\mathbf{x}_1 \neq \mathbf{x}_2 \Longrightarrow \mathbf{A}\mathbf{x}_1 \neq \mathbf{A}\mathbf{x}_2$$

Finding the matrix representation of $\mathbf{A}^{-1}$ is not an easy task however.

- If $\mathbf{A}$ is 1×1 then $\mathbf{A} = a$ and the inverse is $1/a$.
- If $\mathbf{A}$ is 2×2 then

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \text{ and } \mathbf{A}^{-1} = \frac{1}{D}\begin{bmatrix} d & -c \\ -b & a \end{bmatrix} \text{ provided } D = ad - bc \neq 0$$

Some results on inverses are as follows:

1. If $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are matrices such that
$$\mathbf{AB} = \mathbf{CA} = \mathbf{I}$$
then $\mathbf{A}^{-1} = \mathbf{B} = \mathbf{C}$.
2. A matrix $\mathbf{A}$ is invertible if and only if $\mathbf{Ax} = \mathbf{0}$ implies $\mathbf{x} = \mathbf{0}$ or equivalently if and only if every $\mathbf{y}$ can be written as $\mathbf{y} = \mathbf{Ax}$.
3. If $\mathbf{A}$ and $\mathbf{B}$ are invertible so is $\mathbf{AB}$ and $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
4. If $\mathbf{A}$ is invertible and $\alpha \neq 0$ then $\alpha\mathbf{A}$ is invertible and $(\alpha\mathbf{A})^{-1} = \alpha^{-1}\mathbf{A}^{-1}$
5. If $\mathbf{A}$ is invertible then $\mathbf{A}^{-1}$ is invertible and $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$

In many problems one "guesses" the inverse of $\mathbf{A}$ and then verifies that it is in fact the inverse. The following results help.

(1) Let
$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \text{ where } \mathbf{A} \text{ has an inverse}$$
Then
$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\mathbf{Q}^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{Q}^{-1} \\ -\mathbf{Q}^{-1}\mathbf{C}\mathbf{A}^{-1} & \mathbf{Q}^{-1} \end{bmatrix}$$
where $\mathbf{Q} = \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$.

(2) Let
$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \text{ where } \mathbf{D} \text{ has an inverse}$$
Then
$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & -(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & \mathbf{D}^{-1}+\mathbf{D}^{-1}\mathbf{C}(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \end{bmatrix}$$

(3) Let $\mathbf{A}$ be n by n, $\mathbf{U}$ be m by n, $\mathbf{S}$ be m by m and $\mathbf{V}$ be m by n. Then if $\mathbf{A}$, $\mathbf{A}+\mathbf{U}^\top\mathbf{S}\mathbf{V}$ and $\mathbf{S}+\mathbf{S}\mathbf{V}\mathbf{A}^{-1}\mathbf{U}^T\mathbf{S}$ are each invertible we have
$$\left[\mathbf{A}+\mathbf{U}^\top\mathbf{S}\mathbf{V}\right]^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}^\top\mathbf{S}\left[\mathbf{S}+\mathbf{S}\mathbf{V}\mathbf{A}^{-1}\mathbf{U}^\top\mathbf{S}\right]^{-1}\mathbf{S}\mathbf{V}\mathbf{A}^{-1}$$

(4) If $\mathbf{S}^{-1}$ exists in the preceeding result then
$$\left[\mathbf{A}+\mathbf{U}^\top\mathbf{S}\mathbf{V}\right]^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}^\top\left[\mathbf{S}^{-1}+\mathbf{V}\mathbf{A}^{-1}\mathbf{U}^T\right]^{-1}\mathbf{V}\mathbf{A}^{-1}$$

(5) If $\mathbf{S}=\mathbf{I}$ and $\mathbf{u}$ and $\mathbf{v}$ are vectors then
$$\left[\mathbf{A}+\mathbf{u}\mathbf{v}^\top\right]^{-1} = \mathbf{A}^{-1} - \frac{1}{(1+\mathbf{v}^\top\mathbf{A}^{-1}\mathbf{u})}\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^\top\mathbf{A}^{-1}$$

7.1.7 Determinants

Definition 7.8 The *determinant* $\det(\mathbf{A})$ of a p by p square matrix $\mathbf{A}=(a_{ij})$, is defined as
$$\sum_\pi sgn(\pi) a_{\pi(1)1}a_{\pi(2)2}\cdots a_{\pi(p)p}$$
where

- π is a *permutation* of the integers $1,2,\ldots,p$ (a permutation of a set is an ordering of the set e.g. (2,1) is a permutation of $\{1,2\}$ as is (1,2). (There are $p!$ permutations of the integers $\{1,2,\ldots,p\}$)
- $sgn(\pi)=1$ if the number of transpositions needed to change
$$(1,2,\ldots,p) \text{ into } (\pi(1),\pi(2),\ldots,\pi(p))$$
is even and $sgn(\pi)=-1$ if the number of transpositions needed is odd. (A *transposition* consists of interchanging two of the coordinates in $(1,2,\ldots,p)$).

example:

$$\text{If } \mathbf{A} = \begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix}$$

then $p = 2$ and $\{1, 2\}$ has two permutations $(1, 2)$ and $(2, 1)$ so that $sgn((1, 2)) = 1$ and $sgn((2, 1)) = -1$. Thus $\det(\mathbf{A}) = (1)a_{11}a_{22} - (-1)a_{12}a_{21} = 3 \times 2 - 1 \times 2 = 4$.

Properties of determinants

- $\det(\mathbf{AB}) = \det(A)\det(B)$
- $\det(\mathbf{A}^\top) = \det(\mathbf{A})$
- $\det(a\mathbf{A}) = a^p \det(\mathbf{A})$ for any real number a

7.1.8 Minors, Cofactors and Adjoints

Definition 7.9 A *submatrix* of a matrix **A** is any matrix obtained by deleting rows and/or columns of **A**.

Definition 7.10 The *minor* of an element a_{ij} of the square matrix **A** is the determinant of the submatrix of **A** defined by deleting the ith row and jth column of **A**.

Definition 7.11 The *cofactor* of a_{ij}, denoted by C_{ij} is defined as

$$C_{ij} = (-1)^{i+j}(\text{minor of } a_{ij})$$

If **C** is the matrix of cofactors of **A**, $\mathbf{C}^\top$ is called the *adjoint* of **A**.

example:

$$\text{If } \mathbf{A} = \begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix} \text{ then } \mathbf{C}^\top = adjoint(\mathbf{A}) = \begin{bmatrix} 2 & -1 \\ -3 & 2 \end{bmatrix}$$

Since

$$\begin{aligned} \text{minor of } a_{11} &= |2| \ ; \ C_{11} = 2 \\ \text{minor of } a_{12} &= |3| \ ; \ C_{12} = -3 \\ \text{minor of } a_{21} &= |1| \ ; \ C_{21} = -1 \\ \text{minor of } a_{22} &= |2| \ ; \ C_{22} = 2 \end{aligned}$$

Properties of cofactors and adjoints

- $\det(\mathbf{A}) = \sum_{i=1}^{p} a_{ij}C_{ij} = \sum_{j=1}^{p} a_{ij}C_{ij}$
- $\mathbf{A}^{-1} = [adjoint(\mathbf{A})]/\det(\mathbf{A})$

examples:

$$\begin{aligned} det(\mathbf{A}) &= 2(2) + 1(-3) = 1 \\ det(\mathbf{A}) &= 3(-1) + 2(2) = 1 \\ det(\mathbf{A}) &= 2(2) + 3(-1) = 1 \\ det(\mathbf{A}) &= 1(3) + 2(2) = 1 \end{aligned}$$

$$\text{If } \mathbf{A} = \begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix} \text{ then } \mathbf{A}^{-1} = \begin{bmatrix} 2 & -1 \\ -3 & 2 \end{bmatrix}$$

Using determinants thus provides us with a method of finding the inverse of a matrix.

7.2 Vector Spaces

7.2.1 Definition and Examples

Definition 7.12 A vector space $\mathcal{V}$ is a set of points (called vectors) satisfying the following conditions:

(1) An operation + exists which satisfies the following properties:

(a) $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$

(b) $\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$

(c) A vector $\mathbf{0}$ exists in $\mathcal{V}$ such that $\mathbf{x} + \mathbf{0} = \mathbf{x}$ for every $\mathbf{x} \in \mathcal{V}$

(d) For every $\mathbf{x} \in \mathcal{V}$ a vector $-\mathbf{x}$ exists in $\mathcal{V}$ such that $-\mathbf{x} + \mathbf{x} = \mathbf{0}$

(2) An operation $\cdot$ exists which satisfies the following properties:

(a) $\alpha \cdot (\mathbf{x} + \mathbf{y}) = \alpha \cdot \mathbf{y} + \alpha \cdot \mathbf{x}$

(b) $\alpha \cdot (\beta \cdot \mathbf{x}) = (\alpha\beta) \cdot \mathbf{x}$

(c) $(\alpha + \beta) \cdot \mathbf{x} = \alpha \cdot \mathbf{x} + \beta \cdot \mathbf{x}$

(d) $1 \cdot \mathbf{x} = \mathbf{x}$

where the scalars α and β are real numbers. For ease of notation we shall eliminate the $\cdot$ for scalar multiplication.

The $\mathbf{0}$ vector will be called the **null vector** or the **origin**.

example 1: Let $\mathbf{x}$ represent a point in two dimensional space with addition and scalar multiplication defined by

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \end{bmatrix} \text{ and } \alpha \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \alpha\, x_1 \\ \alpha\, x_2 \end{bmatrix}$$

The origin and negatives are defined by

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ and } - \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -x_1 \\ -x_2 \end{bmatrix}$$

example 2: Let $\mathbf{x}$ represent a point in n dimensional space (called Euclidean space and denoted by $\mathbf{R}^n$) with addition and scalar multiplication defined by

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix} \text{ and } \alpha \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \alpha\, x_1 \\ \alpha\, x_2 \\ \vdots \\ \alpha\, x_n \end{bmatrix}$$

The origin and negatives in this case are defined by

$$\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \text{ and } - \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} -x_1 \\ -x_2 \\ \vdots \\ -x_n \end{bmatrix}$$

7.2.2 Linear Independence and Bases

Definition 7.13 A finite set of vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ is said to be a **linearly independent set** if

$$\sum_i \alpha_i \mathbf{x}_i = \mathbf{0} \Longrightarrow \alpha_i = 0 \text{ for each } i$$

If a set of vectors is not linearly independent it is said to be **linearly dependent**.

If the set of vectors is empty we define $\sum_i \mathbf{x}_i = \mathbf{0}$ so that, by convention, the empty set of vectors is a linearly independent set of vectors.

Results on linear independence:

- The set $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ is linearly dependent if and only if

$$\mathbf{x}_t = \sum_{i=1}^{t-1} \alpha_i \mathbf{x}_i \text{ for some } t \geq 2$$

- A **linear basis** or **coordinate system** in a vector space $\mathcal{V}$ is a set $\mathcal{E}$ of linearly independent vectors in $\mathcal{V}$ such that each vector in $\mathcal{V}$ can be written as a linear combination of the vectors in $\mathcal{E}$.
- Since the vectors in $\mathcal{E}$ are linearly independent the representation as a linear combination is unique. If the number of vectors in $\mathcal{E}$ is finite we say that the vector space $\mathcal{V}$ is **finite dimensional**.
- The **dimension** of a vector space is the number of vectors in any basis of the vector space.
- Every vector space has a basis.

In most applications we will use the canonical basis for $\mathbf{R}^n$ defined by

$$\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$$

where

$$\mathbf{e}_j = (e_{kj}) = \begin{cases} 1 & k = j \\ 0 & k \neq j \end{cases}$$

i.e. $\mathbf{e}_j$ is $n \times 1$ with j row equal to 1 and 0 in every other row.

Definition 7.14 A non-empty subset $\mathcal{M}$ of a vector space $\mathcal{V}$ is called a **subspace** if $\mathbf{x}, \mathbf{y} \in \mathcal{V}$ implies that every linear combination $\alpha\mathbf{x} + \beta\mathbf{y} \in \mathcal{M}$.

Theorem 7.1 If $\{\mathbf{x}_i, i \in I\}$ is a set of vectors the subspace **spanned** by $\{\mathbf{x}_i, i \in I\}$ is the set of all linear combinations of the vectors in $\{\mathbf{x}_i, i \in I\}$ and is denoted by $\mathbf{sp}\left(\{\mathbf{x}_i, i \in I\}\right)$, It follows that an alternative characterization of a basis is that it is a set of linearly independent vectors which spans $\mathcal{V}$.

example 1: If $\mathcal{X}$ is the empty set then the space spanned by $\mathcal{X}$ is the $\mathbf{0}$ vector.
example 2: In $\mathbf{R}^3$ let $\mathcal{X}$ be the space spanned by $\mathbf{e}_1, \mathbf{e}_2$ where

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \text{ and } \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

Then $\mathcal{X}$ is the set of all vectors of the form

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix}$$

7.2.3 Geometry

Definition: In $\mathbf{R}^n$ the **inner product** of $\mathbf{x}$ and $\mathbf{y}$ is defined as

$$\mathbf{x}^\top \mathbf{y}$$

Definition: In $\mathbf{R}^n$ the **length** of $\mathbf{x}$ is defined as

$$\sqrt{\mathbf{x}^\top \mathbf{x}}$$

Definition: In $\mathbf{R}^n$ two vectors $\mathbf{x}$ and $\mathbf{y}$ are **orthogonal** if their inner product is 0 i.e. if

$$\mathbf{x}^\top \mathbf{y} = 0$$

Definition: More generally in $\mathbf{R}^n$ the cosine of the angle θ between two vectors $\mathbf{x}$ and $\mathbf{y}$ is defined by

$$\cos\theta = \frac{\mathbf{x}^\top \mathbf{y}}{\sqrt{\mathbf{x}^T \mathbf{x} \mathbf{y}^T \mathbf{y}}}$$

Definition: In $\mathbf{R}^n$ the **distance** between two vectors $\mathbf{x}$ and $\mathbf{y}$ is defined by

$$\sqrt{(\mathbf{x} - \mathbf{y})^\top (\mathbf{x} - \mathbf{y})}$$

i.e. as the length of $\mathbf{x} - \mathbf{y}$.

Definition: In $\mathbf{R}^n$ the **distance** between a vector $\mathbf{x}$ and a subspace $\mathcal{Y}$ is defined as

$$\min_{\mathbf{y} \in \mathcal{Y}} \sqrt{(\mathbf{x} - \mathbf{y})^\top (\mathbf{x} - \mathbf{y})}$$

Definition: $\mathcal{X}$ is said to be an **orthonormal** set of vectors if

$$(\mathbf{x}, \mathbf{y}) = \begin{cases} 1 & \text{if } \mathbf{x} = \mathbf{y} \\ 0 & \text{if } \mathbf{x} \neq \mathbf{y} \end{cases}$$

$\mathbf{X}$ is said to be a **complete orthonormal** set of vectors if it is not contained in a larger set of orthonormal vectors.

7.2.4 Basic Results

Result: If $\mathcal{X}$ is an orthonormal set then its vectors are linearly independent.

Result: (Cauchy Schwartz Inequality) If $\mathbf{x}$ and $\mathbf{y}$ are vectors in an inner product space then

$$|(\mathbf{x}, \mathbf{y})| \leq ||\mathbf{x}|| \, ||\mathbf{y}||$$

Result: If $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ is any finite orthonormal set in a vector space $\mathcal{V}$ then the following conditions are equivalent:

(1) $\mathcal{X}$ is complete.

(2) $(\mathbf{x}, \mathbf{x}_i) = \mathbf{0}$ for $i = 1, 2, \ldots, n$ implies that $\mathbf{x} = \mathbf{0}$.

(3) The space spanned by $\mathcal{X}$ is equal to $\mathcal{V}$.

(4) If $\mathbf{x} \in \mathcal{V}$ then $\mathbf{x} = \sum_i (\mathbf{x}, \mathbf{x}_i)\mathbf{x}_i$.

(5) If $\mathbf{x}$ and $\mathbf{y}$ are in $\mathcal{V}$ then

$$(\mathbf{x}, \mathbf{y}) = \sum_i (\mathbf{x}, \mathbf{x}_i)(\mathbf{x}_i, \mathbf{y})$$

(6) If $\mathbf{x} \in \mathcal{V}$ then

$$||\mathbf{x}||^2 = \sum_i |(\mathbf{x}_i, \mathbf{x})|^2$$

7.2.5 Gram Schmidt Process

The Gram Schmidt process can be used to construct an orthonormal basis for a vector space. Start with a basis for $\mathcal{V}$ as $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ and form

$$\mathbf{y}_1 = \frac{\mathbf{x}_1}{||\mathbf{x}_1||}$$

Next define

$$\mathbf{z}_2 = \mathbf{x}_2 - (\mathbf{x}_2, \mathbf{y}_1)\mathbf{y}_1$$

Since the $\mathbf{x}_i$ are linearly independent $\mathbf{z}_2 \neq \mathbf{0}$ and $\mathbf{z}_2$ is orthogonal to $\mathbf{y}_1$. Hence $\mathbf{y}_2 = \frac{\mathbf{z}_2}{||\mathbf{z}_2||}$ is orthogonal to $\mathbf{y}_1$ and has unit length. If $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_r$ have been so chosen then we form

$$z_{r+1} = \mathbf{x}_{r+1} - \sum_{i=1}^{r} (\mathbf{x}_{r+1}, \mathbf{y}_i)\mathbf{y}_i$$

Since the $\mathbf{x}_i$ are linearly independent $||\mathbf{z}_{r+1}|| > 0$ and since $\mathbf{z}_{r+1} \perp \mathbf{y}_i$ for $i = 1, 2, \ldots, r$ it follows that

$$\mathbf{y}_{r+1} = \frac{\mathbf{z}_{r+1}}{||\mathbf{z}_{r+1}||}$$

may be "added" to the set $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_r\}$ to form a new orthonormal set. The process necessarily stops with $\mathbf{y}_n$ since there can be at most n elements in a linearly independent set.

7.2.6 Orthogonal Projections

Result: (Orthogonal Projection) Let $\mathcal{U}$ be a subspace of an inner product space $\mathcal{V}$ and let $\mathbf{y}$ be a vector in $\mathcal{V}$ which is not in $\mathcal{U}$. Then there exists a unique vector $\mathbf{y}_{\mathcal{U}} \in \mathcal{U}$ and and a unique vector $\mathbf{e} \in \mathcal{V}$ such that

$$\mathbf{y} = \mathbf{y}_{\mathcal{U}} + \mathbf{e} \;\text{ and }\; \mathbf{e} \perp \mathcal{U}$$

Definition: The vector $\mathbf{e}$ in the previous result is called the **orthogonal projection** from **$\mathbf{y}$ to $\mathcal{U}$** and the vector $\mathbf{y}_{\mathcal{U}}$ is called the **orthogonal projection** of **$\mathbf{y}$ on $\mathcal{U}$**.

Result: The projection of $\mathbf{y}$ on $\mathcal{U}$ has the property that

$$||\mathbf{y} - \mathbf{y}_{\mathcal{U}}|| = \overset{\min}{\mathbf{x}} \; \{||\mathbf{y} - \mathbf{x}|| : \; \mathbf{x} \in \mathcal{U}\}$$

with the minimum occurring when $\mathbf{x} = \mathbf{y}_{\mathcal{U}}$. Thus the projection minimizes the distance from $\mathcal{U}$ to $\mathbf{x}$.

7.3 Matrix Rank and Linear Equations

7.3.1 Rank of a Matrix

The rows and columns of a matrix may be considered as vectors in a vector space of the appropriate dimension.

Definition: If $\mathbf{A}$ is an n by p matrix then

(1) The row rank of $\mathbf{A}$ is the number of linearly independent rows of the matrix considered as vectors in p dimensional space.

(2) The column rank of $\mathbf{A}$ is the number of linearly independent columns of the matrix considered as vectors in n dimensional space.

Result: row rank of $\mathbf{A}$ = column rank of $\mathbf{A}$

We thus define the rank of a matrix $\mathbf{A}$, $\rho(\mathbf{A})$ to be the number of linearly independent rows or the number of linearly independent columns in the matrix $\mathbf{A}$. Note that $\rho(\mathbf{A})$ is unaffected by pre or post multiplication by non singular matrices.

Results on ranks of matrices:

- An n by n matrix is non singular if and only if it is of rank n.
- If $\mathbf{A}$ is an n by p matrix then rank $\mathbf{A}^\top\mathbf{A}$ = rank $\mathbf{A}$ = rank $\mathbf{A}\mathbf{A}^T$

Definition: The **trace** of a square matrix is the sum of its diagonal elements i.e.

$$tr(\mathbf{A}) = \sum_i a_{ii}$$

- $tr(\mathbf{AB}) = tr(\mathbf{BA})$; $tr(\mathbf{A}+\mathbf{B}) = tr(\mathbf{A}) + tr(\mathbf{B})$; $tr(\mathbf{A}^\top) = tr(\mathbf{A})$
- If $\mathbf{A}$ is idempotent ($\mathbf{A}^2 = \mathbf{A}$) then

$$\text{rank } \mathbf{A} = tr(\mathbf{A})$$

7.4 Linear Equations

Consider a set of n linear equations in p unknowns

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1p}x_p & = & y_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2p}x_p & = & y_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1}x_1 & + & a_{n2}x_2 & + & \cdots & + & a_{np}x_p & = & y_n \end{array}$$

These equations can be written compactly in matrix notation as

$$\mathbf{A}\mathbf{x} = \mathbf{y}$$

where $\mathbf{A}$ is the n by p matrix with i, j element equal to a_{ij}, $\mathbf{x}$ is the p by one column vector with jth element equal to x_j and $\mathbf{y}$ is the n by one column vector with ith element equal to y_i.

Often such "equations" arise without knowledge of whether they are really equations, i.e. does there exist a vector $\mathbf{x}$ which satisfies the equations? If such an $\mathbf{x}$ exists the equations are said to be **consistent**, otherwise they are said to be **inconsistent**.

Result: The equations $\mathbf{A}\mathbf{x} = \mathbf{y}$ are consistent if and only if

$$\text{rank}([\mathbf{A}, \mathbf{y}]) = \text{rank}(\mathbf{A})$$

In most situations $n = p$ so that we have p equations in p unknowns. In this case we have the following results:

1. If the rank of $\mathbf{A}$ is p (i.e. $\mathbf{A}$ has an inverse) then the equations have a unique solution $\mathbf{x}$ given by
$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$$

2. If the rank of $\mathbf{A}$ is less than p and the rank of $[\mathbf{A}, \mathbf{y}]$ is equal to the rank of $\mathbf{A}$ then there are many solutions. One such solution is
$$\mathbf{x} = \mathbf{A}^{-}\mathbf{y}$$
where $\mathbf{A}^{-}$ satisfies
$$\mathbf{A}\mathbf{A}^{-}\mathbf{A} = \mathbf{A}$$
and is called a generalized inverse of $\mathbf{A}$.

For any matrix there a generalized inverse always exists.

example:

$$\text{If } \mathbf{A} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \text{ then a generalized inverse of } \mathbf{A} \text{ is } \mathbf{A}^- = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 2 \end{bmatrix}$$

so that the equations

$$\begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \\ 3 \end{bmatrix}$$

have a solution given by

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}$$

7.5 Characteristic Roots and Vectors

7.5.1 Definitions

Definition: A scalar λ is a **characteristic root** and a non-zero vector $\mathbf{x}$ is a **characteristic vector** of the matrix $\mathbf{A}$ if

$$\mathbf{A}\mathbf{x} = \lambda\,\mathbf{x}$$

Other names for characteristic roots are proper value, latent root, eigenvalue and secular value with similar adjectives appying to characteristic vectors. By convention we standardize characteristic vectors so that

$$\mathbf{x}^\top\mathbf{x} = 1$$

7.5.2 Results and Properties

Result: λ is a characteristic root of $\mathbf{A}$ if and only if $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$.

Result: If $\mathbf{A}$ is a real symmetric matrix with characteristic roots $\lambda_1, \lambda_2, \ldots, \lambda_n$ and characteristic vectors $x_1, x_2, \ldots, x_n$ then

$$\mathbf{A} = \sum_{i=1}^{n} \lambda_i \mathbf{x}_i \mathbf{x}_i^\top$$

This representation of $\mathbf{A}$ is called the **spectral representation** of $\mathbf{A}$.

Definition: If $\mathbf{A}$ is a symmetric matrix with real elements $\mathbf{x}^\top\mathbf{A}\mathbf{x}$ is called a **quadratic form**. A quadratic form is said to be

- **positive definite** if $\mathbf{x}^\top\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$
- **non negative definite** if $\mathbf{x}^\top\mathbf{A}\mathbf{x} \geq 0$ for all $\mathbf{x} \neq \mathbf{0}$

If $\mathbf{A}$ is positve definite then all of its characteristic roots are positive while if $\mathbf{A}$ is non negative definite then all of its characteristic roots are non negative.

- If $\lambda_i \neq \lambda_j$ then the corresponding characteristic vectors are orthogonal
- The characteristic roots of a real symmetric matrix are real.
- $tr(A) = \sum_{i=1}^{p} \lambda_i$
- $\det(A) = \prod_{i=1}^{p} \lambda_i$
- $rank(A) =$ number of non-zero eigenvalues of $A^\top A$ or AA^T

7.5.3 The Singular Value Decomposition (SVD) of a Matrix

If $\mathbf{A}$ is any n by p matrix then there exists matrices $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{D}$ such that

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$$

where

- $\mathbf{U}$ is n by p and $\mathbf{U}^\top\mathbf{U} = \mathbf{I}_p$
- $\mathbf{V}$ is p by p and $\mathbf{V}^\top\mathbf{V} = \mathbf{I}_p$
- $\mathbf{D} = diag(\sigma_1, \sigma_2, \ldots, \sigma_p)$ and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$

7.6 Matrix Calculus

Definition: If $\mathbf{y} = \mathbf{f}(\mathbf{x})$ is a vector of p real valued functions

$$\begin{array}{rcl} y_1 & = & f_1(\mathbf{x}) \\ y_2 & = & f_2(\mathbf{x}) \\ \vdots & \vdots & \vdots \\ y_p & = & f_p(\mathbf{x}) \end{array}$$

where f_i is a function of a $q \times 1$ vector $\mathbf{x}$ then we define the derivative of $\mathbf{y}$ with respect to $\mathbf{x}$ as the $p \times q$ matrix with i, j element equal to $\frac{\partial y_i}{\partial x_j}$ i.e.

$$\frac{d\mathbf{y}}{d\mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_q} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_q} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_p}{\partial x_1} & \frac{\partial y_p}{\partial x_2} & \cdots & \frac{\partial y_p}{\partial x_q} \end{bmatrix}$$

example: Let $\mathbf{y} = \mathbf{a} + \mathbf{B}\mathbf{x}$ where $\mathbf{a}$ and $\mathbf{B}$ do not depend on $\mathbf{x}$. Then

$$\frac{d\mathbf{y}}{d\mathbf{x}} = \mathbf{B}$$

To see this note that

$$y_i = a_i + \sum_{k=1}^{q} b_{ik} x_k$$

so that

$$\frac{\partial y_i}{\partial x_j} = b_{ij}$$

example: If $y = \mathbf{x}^\top \mathbf{x}$ then

$$\frac{d\mathbf{y}}{d\mathbf{x}} = 2\mathbf{x}^\top$$

To see this note that

$$y = \sum_{i=1}^{q} x_i^2$$

so that

$$\begin{aligned}\frac{d\mathbf{y}}{d\mathbf{x}} &= \left[\frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2}, \cdots \frac{\partial y}{\partial x_q}\right] \\ &= [2x_1, 2x_2, \ldots, 2x_q] \\ &= 2\mathbf{x}^\top\end{aligned}$$

Theorem: If $\mathbf{y} = \mathbf{f}(\mathbf{u})$ is a $p \times 1$ vector of functions of a $q \times 1$ vector $\mathbf{u}$, where $\mathbf{u}$ is a function of a $r \times 1$ vector $\mathbf{x}$ then $\frac{d\mathbf{y}}{d\mathbf{x}}$ is the $p \times r$ matrix given by

$$\frac{d\mathbf{y}}{d\mathbf{x}} = \left[\frac{d\mathbf{y}}{d\mathbf{u}}\right]\left[\frac{d\mathbf{u}}{d\mathbf{x}}\right]$$

example: If $s = (\mathbf{y} - \mathbf{X}\mathbf{b})^\top(\mathbf{y} - \mathbf{X}\mathbf{b})$ then

$$\frac{ds}{d\mathbf{b}} = -2(\mathbf{y} - \mathbf{X}\mathbf{b})^\top\mathbf{X}$$

To see this let $\mathbf{u} = \mathbf{y} - \mathbf{X}\mathbf{b}$ then

$$\begin{aligned}\frac{ds}{d\mathbf{b}} &= \left[\frac{ds}{d\mathbf{u}}\right]\left[\frac{du}{d\mathbf{b}}\right] \\ &= 2\mathbf{u}^\top(-\mathbf{X}) \\ &= -2(\mathbf{y} - \mathbf{X}\mathbf{b})^\top\mathbf{X}\end{aligned}$$

example: If $y = \mathbf{x}^\top\mathbf{A}\mathbf{x}$ where $\mathbf{A}$ is a symmetric non-negative matrix then

$$\frac{dy}{d\mathbf{x}} = 2\mathbf{x}^\top\mathbf{A}$$

To see this let $\mathbf{A} = \mathbf{B}\mathbf{B}^\top$ and $\mathbf{u} = \mathbf{B}^T\mathbf{x}$. Then $y = \mathbf{u}^\top\mathbf{u}$ so that

$$\begin{aligned}\frac{dy}{d\mathbf{x}} &= \left[\frac{dy}{d\mathbf{u}}\right]\left[\frac{du}{d\mathbf{x}}\right] \\ &= 2\mathbf{u}^\top\mathbf{B}^T \\ &= 2\mathbf{x}^\top\mathbf{B}\mathbf{B}^T \\ &= 2\mathbf{x}^\top\mathbf{A}\end{aligned}$$

A Complex Numbers

Complex numbers are defined as follows. z is a complex number if it is of the form $z = (x, y)$ where x and y are real numbers and

- $(x, 0)$ equals the real number x
- $(0, 1)$ is called the imaginary unit denoted by i
- $\mathcal{R}(z) = x$ is called the real part of z
- $\mathcal{C}(z) = y$ is called the imaginary part of z
- $(x_1, y_1) = (x_2, y_2)$ if and only if $x_1 = x_2$ and $y_1 = y_2$
- $z_1 + z_2 = (x_1 + x_2, y_1 + y_2)$
- $z_1 \times z_2 = (x_1x_2 - y_1y_2\ , x_1y_2 + x_2y_1)$
- The *conjugate* of z is $\overline{z} = (x, -y)$
- The absolute value or modulus of z is $|z| = (z \times \overline{z})^{1/2} = (x^2 + y^2)^{1/2}$
- $|z_1 + z_2| \leq |z_1| + z_2|$

Complex numbers have two equivalent representations:

- As a sum $z = x + iy$ where $i^2 = -1$ and the usual operations of addition, subtraction and multiplication are followed for expressions of the form $a + b$
- As an exponential. Define $\exp(z) = e^x[\cos(y) + i\ \sin(y)]$ where z is a complex number. Then $y = 0$ gives the usual definition of the exponential function. Note that for θ real we have
$$\exp(i\theta) = \cos(\theta) + i\sin(\theta) \quad \text{and} \quad \exp(-i\theta) = \cos(\theta) - i\sin(\theta)$$
so that
$$\begin{array}{rcl} \cos(\theta) & = & \frac{\exp(i\theta)+\exp(-i\theta)}{2} = \frac{e^{i\theta}+e^{-i\theta}}{2} \\ \sin(\theta) & = & \frac{\exp(i\theta)-\exp(-i\theta)}{2i} = \frac{e^{i\theta}-e^{-i\theta}}{2i} \end{array}$$
and it follows that
$$z = r[cos(\theta) + i\sin(\theta)] = re^{i\theta}$$
where
$$r = \sqrt{x^2 + y^2} \quad \text{and} \quad \tan(\theta) = y/x$$

B Special Topics

B.1 Gaussian Probability Density Function

If

$$f(x) = \frac{\exp(-\frac{x^2}{2})}{\sqrt{2\pi}}$$

Then f is called the standard normal or Gaussian p.d.f.

B.2 Gamma Function

The integral

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

defined for $\alpha > 0$ is called the *Gamma function* with parameter α.

Properties

- $\Gamma(1) = \int_0^\infty e^{-x} dx = 1$
- $\Gamma\left(\frac{1}{2}\right) = \int_0^\infty x^{-\frac{1}{2}} e^{-x} dx = \sqrt{\pi}$
- $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ Thus if n is a positive integer

$$\Gamma(\alpha) = (n-1)!$$

B.3 Beta Function

The quantity

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx$$

is called the Beta function with parameters α and β.

Properties

- $B(\alpha, \beta) = B(\beta, \alpha)$
-
$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

  This fact is established in many advanced calculus texts.

B.4 Stirling's Approximation to $n!$

$$\lim_{n\to\infty}\left[\frac{n!}{\sqrt{2\pi}\, n^{n+\frac{1}{2}}e^{-n}}\right] = 1$$

A more precise version of Stirling's approximation is

$$n! = \sqrt{2\pi}\, n^{n+\frac{1}{2}}e^{-n}e^{\theta_n}$$

where

$$\frac{1}{12n+1} < \theta_n < \frac{1}{12n}$$