Mathematics Notes

Gregory Hill

Friday 1st September, 2017

1 About

Upon completion of my undergraduate at Abertay University, I decided to continue my studies at the University of Edinburgh. I created these notes over the summer of 2017 to help understand and formulate the mathematical pre-requisites of MSc Informatics.

2 Set Theory

Type	Symbol	Example
Natural Numbers	$\mathbb{N}$	$\{1, 2, 3, 4, \ldots\}$
Integers	$\mathbb{Z}$	$\{\ldots, -4, -3, -2, -1, 0, 1, 2, 3, 4, \ldots\}$
Rational Numbers	$\mathbb{Q}$	
Real Numbers	$\mathbb{R}$	π, e
Complex Numbers	$\mathbb{C}$	

All fraction numbers. $x \in \mathbb{Q}$ if x can be written in the form p/q where p and q are integers with $q \neq 0$.

The set of real numbers that are not rational are called irrational.

For $x \in \mathbb{R}$, x is any number with decimal representation (i.e. π, e).
A number $x \in \mathbb{C}$ if x can be expressed in the form $a + bi$, where a and b are real numbers and i is the imaginary unit, satisfying the equation $i^2 = -1$.

The **empty set** is defined as $\emptyset$.

3 Theorem Terminology

3.1 Def ^{n} (Definition), Axiom or Postulate

A **definition** provides some meaning, terminology or structure. It is typically accepted without proof. An **axiom** or **postulate** is a statement that is also accepted without proof and regarded as fundamental to a subject - in modern mathematics they are merely 'background' assumptions we make.

3.2 Proposition / Lemma

Simple facts from the definition, a theorem of lesser importance - often with a simple proof.

3.3 Theorem

The main result. The term is commonly reserved for the most important results. There are three types:

- Existence
- Constructive
- Closed-Form

3.4 Corollary

An immediate consequence of the theorem. A proposition that follows with little proof from another theorem or definition.

4 Calculus

4.1 Differentiation

The rate of change of function $f(x)$ with respect to x (in 2-D where $y = f(x)$) is denoted by $f'(x)$ or $\frac{dy}{dx}$.

$f(x)$	$f'(x)$	$f(x)$	$f'(x)$
x	1	x^n	nx^{n-1}
$\sin x$	$\cos x$	$\cos x$	$-\sin x$
$\tan x$	$\sec^2 x$	$\sec x$	$\sec x \tan x$
$\cot x$	$-\csc^2 x$	$\sin^{-1} x$	$\frac{1}{\sqrt{1-x^2}}$
$\csc x$	$-\csc x \cot x$	$\tan^{-1} x$	$\frac{1}{1+x^2}$
e^x	e^x	$\ln x$	$\frac{1}{x}$

$$\begin{aligned}
\textbf{Linear Rule}: &\quad (cf)'(x) = (c(f'(x)) \quad c \in \mathbb{R} \\
&\quad (f+g)'(x) = f'(x) + g'(x) \\
\textbf{Product Rule}: &\quad \frac{d}{dx}[f(x)g(x)] = f'(x)g(x) + f(x)g'(x) \\
\textbf{Quotient Rule}: &\quad (\frac{f}{g})'(x) = \frac{d}{dx}(\frac{f(x)}{g(x)}) = \frac{f'(x)g(x) - f(x)g'(x)}{[g(x)]^2} \\
\textbf{Reciprocal Rule}: &\quad (\frac{1}{g})'(x) = \frac{d}{dx}(\frac{1}{g(x)}) = -\frac{g'(x)}{[g(x)]^2} \\
\textbf{Chain Rule}: &\quad \frac{d}{dx}[f(g(x))] = f'(g(x))g'(x)
\end{aligned}$$

- An equation $y = f(x)$ defines y **explicitly** as a function of x.
- An equation $f(x, y) = 0$ defines y **implicitly** as a function of x.

4.2 Partial Derivatives

A partial derivate of a function of several variables is its derivate with respect to one of those variables with the others held constant.

For $z = f(x, y)$: $\frac{\partial z}{\partial x} = z_x = \frac{\partial f}{\partial x} = f_x$

High Order:

$$\frac{\partial^2 z}{\partial x^2} = \frac{\partial}{\partial x}(\frac{\partial z}{\partial x}) = z_{xx}$$

$$\frac{\partial^2 z}{\partial y^2} = \frac{\partial}{\partial y}(\frac{\partial z}{\partial y}) = z_{yy}$$

(Mixed):

$$\frac{\partial^2 z}{\partial x \partial y} = \frac{\partial}{\partial x}(\frac{\partial z}{\partial y}) = z_{yx}$$

$$\frac{\partial^2 z}{\partial y \partial x} = \frac{\partial}{\partial y}(\frac{\partial z}{\partial x}) = z_{xy}$$

Suppose $z = f(x(t), y(t))$ where the variables x and y are both functions of t. To calculate the rate z changes with respect to t we use the **chain rule**:

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}$$

An equation $F(x, y, z) = 0$ defines a surface in (x, y, z)-space where z is defined **implicitly** as a function of x and y, therefore we can differentiate implicitly to find $\frac{\partial z}{\partial x}$ and $\frac{\partial z}{\partial y}$.

4.3 Approximation

The **total differential** of z is dz. The values dz, dy and dy represent 'infinitesimal' changes in z, x and y. Given z, we can obtain the partial derivatives and supply the increase or decrease in dx and dy with the original x and y points to calculate Δz:

$$dz = \frac{\partial z}{\partial x} dx + \frac{\partial z}{\partial y} dy$$

Absolute error:

$$\Delta z \approx \frac{\partial z}{\partial x} \Delta x + \frac{\partial z}{\partial y} \Delta y$$

4.4 Stationary Points

For a function of a single variable, $y = f(x)$, stationary points occur where $f'(x) = 0$ and are classified into maxima, minima or points of inflection:

- maxima if $f''(x) < 0$
- minima if $f''(x) > 0$
- point of inflection if $f''(x) = 0$ and changes sign at the stationary point

For a function of two variables, the point (a, b) is a **Stationary Point (or Critical Point)** for the surface $z = f(x, y)$ if:

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = 0$$

We consider the determinant of the **Hessian Matrix** (H) to classify each stationary point $x = a, y = b$.

$$det(H) = f_{xx} f_{yy} - f_{xy}^2$$

- if $det(H) > 0$ and $f_{xx} > 0$ or $f_{yy} > 0$, then **local minimum**.
- if $det(H) > 0$ and $f_{xx} < 0$ or $f_{yy} < 0$, then **local maximum**.
- if $det(H) < 0$, then **saddle point**.
- if $det(H) = 0$, then test fails.

4.5 Integration

Integration is the inverse process to differentiation:

$$\int ax^n dx = \frac{ax^{n+1}}{n+1} + C \quad \textit{provided } n \neq -1$$

Each 'standard derivative' result can be reversed to give the 'standard integral'.

$$\begin{aligned}
\frac{d}{dx}(x) = 1 &\qquad \int 1dx = x + C \\
\frac{d}{dx}(x^n) = nv^{n-1} &\qquad \int x^n dx = \frac{x^{n+1}}{n+1} + C \\
\frac{d}{dx}(\ln|x|) = \frac{1}{x} &\qquad \int \frac{1}{x} dx = \ln|x| + C \\
\frac{d}{dx}(\cos x) = -\sin x &\qquad \int \sin x dx = -\cos x + C \\
\frac{d}{dx}(\sin x) = \cos x &\qquad \int \cos x dx = \sin x + C \\
\frac{d}{dx}(\tan x) = \sec^2 x &\qquad \int \sec^2 x dx = \tan x + C \\
\frac{d}{dx}(e^{ax}) = ae^{ax} &\qquad \int e^{ax} dx = \frac{1}{a} e^{ax} + C
\end{aligned}$$

Where C is the *constant of integration* when it is **indefinite** (that is, without limits).

Linear Rule :

$$\int \alpha f(x) dx = \alpha \int f(x) dx$$

$$\int (f_1(x) + f_2(x)) dx = \int f_1(x) dx + \int f_2(x) dx$$

The **definite integral** of the *integrand* $f(x)$ between the lower limit $x = a$ and the upper limit $x = b$ can be defined as:

$$I = \int_a^b f(x)dx$$

such that the constant terms would cancel each other out when the integral's lower limit is taken away from its upper. This is used to measure the area under the plot of a function.

4.5.1 Substitution

If the integral can be setup such that:

$$\int f(g(x))g'(x)\ dx$$

With both $g(x)$ and its derivative $g'(x)$, then we can substitute in u,

$$\int f(u)\ du$$

Where $u = g(x)$ so $\frac{du}{dx} = g'(x)$ and $du = g'(x)dx$.

4.5.2 Parts

This rule is available for integrating products of two functions:

$$\int u\frac{dv}{dx} = uv - \int v\frac{du}{dx}\ dx$$

It is derived from the product rule of differentiation to simplify integration.

4.6 Double and Repeated Integrals

Given a function $f(x, y)$ we can integrate with respect to one variable, and hold the other constant.

$$\int_{y=g(x)}^{y=h(x)} f(x, y) dy = F(x)$$

$$\int_{x=a}^{x=b} F(x) dx = \int_{b}^{a} \int_{g(x)}^{h(x)} f(x, y)\, dy\, dx$$

The limits of this **repeated integral** define a **region (or field)** of integration in the xy-plane.

To sketch the region of integration, take the values a, b, c and d where $a \leq x \leq b$ and $c \leq y \leq d$ and plot for x and y as boundaries, then sketch inward.

To change the order, hold the inner variable of integration as a constant on the graph and calculate the new limits.

5 Vectors

Vectors can be represented by a **directed line segment**, or by an ordered collection of scalars called its components - in this case we refer to it as a **constant vector**. In 3-D space, vectors have three components. For example, vector $\mathbf{a} = (a_1, a_2, a_3)$.

The **magnitude** of $\mathbf{a}$ is written as $|\mathbf{a}| = \sqrt{a_1^2 + a_2^2 + a_3^2}$. If $|\mathbf{a}| = 1$ then it is a **unit vector**. The three main unit vectors along the axes are: $\mathbf{i} = (1, 0, 0)$, $\mathbf{j} = (0, 1, 0)$ and $\mathbf{k} = (0, 0, 1)$. So, $(a_1, a_2, a_3) = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$. To find the unit vector in the same direction as $\mathbf{a}$, $\mathbf{u_a} = (\frac{a_1}{|\mathbf{a}|}, \frac{a_2}{|\mathbf{a}|}, \frac{a_3}{|\mathbf{a}|})$.

To multiply a vector by a scalar, $\lambda\mathbf{a} = (\lambda a_1, \lambda a_2, \lambda a_3)$. The addition or subtraction of two vectors is componentwise - where $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$ and $\mathbf{a} - \mathbf{b} = \mathbf{a} + (-\mathbf{b})$.

The **scalar (dot) product** of two vectors is $\mathbf{a} \cdot \mathbf{b} = a_1b_1 + a_2b_2 + a_3b_3$. Where the angles are perpendicular if $\mathbf{a} \cdot \mathbf{b} = 0$.

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos\theta$$

The **vector (cross) product** of two vectors is given by:

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} = \mathbf{i} \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} - \mathbf{j} \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} + \mathbf{k} \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$$

Where,

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

$$\mathbf{a} \times \mathbf{b} = |\mathbf{a}||\mathbf{b}| \sin\theta \hat{\mathbf{n}}$$

Vectors whose components are functions of some independent variables (e.g. x, y or z, or time t) are called **vector functions**.

$$\mathbf{u}(t) = u_1(t)\mathbf{i} + u_2(t)\mathbf{j} + u_3(t)\mathbf{k} = (u_1(t), u_2(t), u_3(t))$$

These can be differentiated componentwise:

$$\frac{d\mathbf{u}}{dt} = \dot{\mathbf{u}}(t) = \dot{u_1}(t)\mathbf{i} + \dot{u_2}(t)\mathbf{j} + \dot{u_3}(t)\mathbf{k}$$

Higher derivatives are calculated similarly:

$$\frac{d^2\mathbf{u}(t)}{dt^2} = \ddot{\mathbf{u}}(t) = \ddot{u_1}(t)\mathbf{i} + \ddot{u_2}(t)\mathbf{j} + \ddot{u_3}(t)\mathbf{k}$$

Given some position vector $\mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k}$:

- the velocity vector $\mathbf{v}(t) = \dot{\mathbf{r}}(t)$.
- the speed is $|\mathbf{v}(t)|$.
- the acceleration vector $\mathbf{a}(t) = \dot{\mathbf{v}}(t) = \ddot{\mathbf{r}}(t)$.

A **scalar field** $\phi(x, y, z)$ defines a scalar value ϕ at each point (x, y, z). So, at any point $P(a, b, c)$ in space we can calculate $\phi(a, b, c)$. The equation $\phi(x, y, z) = C$ (where C is a constant) defines a surface S_C comprising all points at which ϕ has the value C.

A **vector field** $\mathbf{F}(x, y, z)$ defines a vector at each point:

$$\mathbf{F}(x, y, z) = F_1(x, y, z)\mathbf{i} + F_2(x, y, z)\mathbf{j} + F_3(x, y, z)\mathbf{k}$$

The **gradient** of a scalar field ϕ is the **vector field**:

$$grad\phi = \nabla\phi = \frac{\partial\phi}{\partial x}\mathbf{i} + \frac{\partial\phi}{\partial y}\mathbf{j} + \frac{\partial\phi}{\partial z}\mathbf{k}$$

The **directional derivative** of a scalar function in the direction of a vector is:

$$D_{\mathbf{u}}\phi = \hat{\mathbf{u}} \cdot \nabla\phi, \quad where \quad \hat{\mathbf{u}} = \frac{\mathbf{u}}{|\mathbf{u}|}$$

The **divergence** (stretching) of the vector field $\mathbf{u}(x, y, z) = u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k}$ is the scalar field:

$$div\ \mathbf{u} = \nabla \cdot \mathbf{u} = \frac{\partial u_1}{\partial x} + \frac{\partial u_2}{\partial y} + \frac{\partial u_3}{\partial z}$$

The **curl** (rotation) of the vector field $\mathbf{u}(x, y, z) = u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k}$ is the vector field:

$$curl\ \mathbf{u} = \nabla \times \mathbf{u} = (\frac{\partial u_3}{\partial y} - \frac{\partial u_2}{\partial z})\mathbf{i} + (\frac{\partial u_3}{\partial x} - \frac{\partial u_1}{\partial z})\mathbf{j} + (\frac{\partial u_2}{\partial x} - \frac{\partial u_1}{\partial y})\mathbf{k}$$

$$= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ u_1 & u_2 & u_3 \end{vmatrix}$$

where the above "determinant" is expanded along the top row.

6 Matrices and Linear Systems

A **matrix** is a rectangular array of real numbers a_{ij} called elements, of size m (*rows*) $\times$ n (*columns*):

$$\begin{pmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & x_{m3} & \dots & x_{mn} \end{pmatrix}$$

Element a_{ij} is in row i and column j. This can be written as $A = [a_{ij}]$.

Definitions:

square	equal number of rows to columns ($m = n$)
zero matrix	0_{mn} is the $m \times n$ matrix whose entries are all zero
	$0_{23} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$
transpose	$m \times n$ matrix A is the $n \times m$ matrix A^T
row vector	if $m = 1$: $\mathbf{y} = (y_1, y_2, \dots, y_n)$
column vector	if $n = 1$: $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}$

The following definitions only apply to square matrices:

diagonal	elements from top left to bottom right, only true if off-diagonal entries are zero
identity matrix	$I_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$
upper triangular	all elements below diagonal are zero
lower triangular	all elements above the diagonal are zero
symmetric	$A^T = A$
skew symmetric	$A^T = -A$

6.1 Matrix Algebra

For any scalar value k, kA is the matrix with elements $[ka_{ij}]$.

Two matrices $A = [a_{ij}]$ and $B = [b_{ij}]$ are **equal** if and only if they are the **same size** and $a_{ij} = b_{ij}$ for each i and j.

If two matrices are the same size then they can be added / subtracted elementwise.

6.2 Matrix Multiplication

The **inner product** of a row vector r and column vector c (with the same number of elements) is the **scalar** quantity.

The product C of two matrices A and B is defined as:

$$c_{ij} = a_{ik}b_{kj}$$

For each i^{th} row of A, and j^{th} column of B. That is,

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{ik}b_{kj}$$

It is essential that the number of columns in matrix A equals the number of rows in matrix B, so that the matrix product $A(m \times n)B(n \times p) = C(m \times p)$.

6.3 Determinants

The definition of a **determinant** is only defined for square matrices. For matrix A, this is given by:

$$det\mathbf{A} = |\mathbf{A}| = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix}$$

Using **Laplace expansion**, expand along any row or column. The determinant is given by $ad - bc$. For example, the minor M_{23} is described by:

$$\begin{vmatrix} A_{11} & A_{12} \\ A_{31} & A_{32} \end{vmatrix}$$

Each of the three minors should be multiplied by the sign in the corresponding position of the ghost matrix and the element itself.

$$\begin{pmatrix} + & - & + \\ - & + & - \\ + & - & + \end{pmatrix}$$

Therefore,

$$|\mathbf{A}| = A_{11} \begin{vmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{vmatrix} - A_{12} \begin{vmatrix} A_{21} & A_{23} \\ A_{31} & A_{33} \end{vmatrix} + A_{13} \begin{vmatrix} A_{21} & A_{22} \\ A_{31} & A_{32} \end{vmatrix}$$

6.4 Inverse

The inverse of a 2×2 matrix is denoted by:

$$A^{-1} = \frac{C^T}{|A|} = \frac{1}{A_{11}A_{22} - A_{12}A_{21}} \begin{pmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{pmatrix}$$

To find the inverse of a 3×3 matrix, find $|A|$ and the matrix of cofactors C. Then, $A^{-1} = \frac{C^T}{|A|}$.

6.5 Eigenvalues & Eigenvectors

Let A be a square matrix. An **eigenvector** of A is a nonzero vector $\boldsymbol{x}$ such that $A\boldsymbol{x}$ is parallel to $\boldsymbol{x}$. λ is an **eigenvalue** of A if there is a vector $\boldsymbol{x} \neq \mathbf{0}$ with $A\boldsymbol{x} = \lambda\boldsymbol{x}$.

Because $\boldsymbol{x} = I\boldsymbol{x}$, $A\boldsymbol{x} = \lambda I\boldsymbol{x}$, which can be rearranged as $(A - \lambda I)\boldsymbol{x} = 0$.

The eigenvalues of A are the scalars λ that satisfy the **characteristic equation**, $\det(A - \lambda I) = 0$. Their eigenvectors can be calculated by solving $(A - \lambda I)\boldsymbol{x} = 0$.

More specifically, $A\boldsymbol{x} = \lambda\boldsymbol{x}$ so,

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \lambda x \\ \lambda y \end{pmatrix}$$

$$a_{11}x + a_{12}y = \lambda x$$
$$a_{21}x + a_{22}y = \lambda y$$

7 Combinatronics

Combinatronics is the study of how to count things. A **permutation** of some objects is essentially an arrangement of them. A **combination** of some objects is equivalent to a selection.

If we use the notation where P_N stands for the number of permutations of N objects, then we have $P_1 = 1$. For N distinct values, there are $N!$ ways of arranging them - i.e. $A, B, C, D = 4 \times 3 \times 2 \times 1 = 24 = 4!$. Thus, $P_N = N!$. The number of possible outcomes when the order matters is $= N^n$.

The number of permutations of r distinguishable objects out of n objects is given by:

$$^nP_r = \frac{n!}{(n-r)!}$$

Permutations of repeated or unnecessary values should be taken out. For $AAABB$, there are 5! ways of arranging the letters including '3!' A's and '2!' B's. Hence, $\frac{5!}{3!2!}$.

For n objects, r of which are distinguishable, a combination is given by:

$$^nC_r = \frac{n!}{(n-r)!r!}$$

8 Probability

Probabilities are numbers between 0 and 1 that give a measure of how likely an event is to occur. The set of possible outcomes for a trial is called the **sample space**. Any subset of a sample space is an **event**.

Notation:

- Letters $(A, B, C, \ldots)$ for events.
- $P(A), P(B), P(C)$ for the probability of those events occuring.
- $P(\overline{A}), P(\overline{B}), P(\overline{C})$ for the probability of those events not occuring.

If a sample space S consists of equally likely outcomes, then the probability of X is defined by:

$$P(X) = \frac{n(X)}{n(S)}$$

Two events are **mutually exclusive** if one occuring means that the other cannot occur. For two independent events A and B, $P(A \text{ or } B) = P(A) + P(B)$. The probability of both events occuring is $P(A \text{ and } B) = P(A) \times P(B)$.

The events A and $\overline{A}$ are called **complimentary** since exactly one of either must occur. Hence,

$$P(A) + P(\overline{A}) = 1$$

$$P(\overline{A}) = 1 - P(A)$$

A **conditional probability** is the probability that one event happens given that another event has already happened. So, P(A is true, given that B is true) = P(A given B) = $P(A|B)$.

Basic addition rule:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Basic multiplication rule:

$$P(A \cap B) = P(A) \times P(B|A)$$

Hence, to solve basic conditions:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Baye's Theorem:

If,

$$P(B \cap A) = P(A)P(B|A) \quad (1)$$

Also,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (2)$$

Substituting (1) into (2) gives:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Thus $P(A|B)$ can be found using $P(B|A)$.

If $P(A|B) = P(A)$ then the events are independent.

9 Random Variables

Captital letters are used to name random variables, while **lower case letters** indicate particular values.

9.1 Discrete

If X is a variable quantity which takes n discrete values $x_1, x_2, \ldots, x_n$ with probabilities $p_1, p_2, \ldots, p_n$ then X is called a **discrete** random variable if $p_1 + p_2 + \ldots + p_n = 1$.

To prove this:

$$\sum_{x=0}^{n} P(X = x) = 1$$

Expectation of X:

$$E(X) = \sum_{x \in X} xp(x)$$

$$E(X) = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n$$

For a variable X, the expected value of an arbitrary function is given by:

$$E[g(x)] = \sum_{x} g(x) f(x)$$

Variation of X:

$$Var(X) = \sigma_X^2 = E(X^2) - (E(X))^2$$

The **standard deviation** is the square root of the variance: $\sigma = \sqrt{Var(X)}$.

9.1.1 Binomial Distribution

For an experiment with n trials, each of which has two mutually exclusive outcomes (success/failure), where the probability (p) remains constant, then $X \sim B(n, p)$.

$$P(X = r) = \binom{n}{r} p^r (1-p)^{n-r}$$

$$E(X) = np \qquad Var(X) = np(1-p)$$

9.1.2 Poisson Distribution

Suited for random events occuring in time or space where μ is the average number of occurences, $X \sim P(\mu)$.

$$P(X = r) = \frac{e^{-\mu}\mu^r}{r!}$$

$$E(X) = \mu \qquad Var(X) = \mu$$

9.2 Continuous

Used to model data sets which are continous in nature, typically grouped between two limits l_1 and l_2 - which can be infinite. They are represented graphically by continuous curves (where the areas under the curve corresponds to probabilities).

$$\int_{l_1}^{l_2} f(x)dx = 1$$

Probability Density Function (PDF):

$$P(x < X \leq x + dx) = f(x)dx$$

Where,

$$P(a < X \leq b) = \int_a^b f(x)d(x)$$

Cumulative Distribution Function (CDF):

$$F(x) = P(X \leq x) = \int_{l_1}^{x} f(x)dx$$

The median m satisfies $P(X \leq m) = \frac{1}{2}$, so $F(m) = \frac{1}{2}$.

The mean can be defined such that:

$$E(X) = \int_{l_1}^{l_2} x f(x) dx$$

So,

$$E(X^2) = \int_{l_1}^{l_2} x^2 f(x) dx$$

Which can both be used to calculate the variance:

$$Var(X) = \sigma_X^2 = E(X^2) - (E(X))^2$$

9.2.1 Normal Distribution

When data tends to a central value with no bias left or right, like a bell-shaped curve, it gets close to a **normal distribution** - symmetrical about the center. Defined by $X \sim N(\mu, \sigma^2)$ for mean μ and standard deviation σ.

The **standardised normal distribution** Z is the normal distribution where $Z \sim N(0, 1)$. For any particular value $X = x$, we can obtain the corresponding value $z = \frac{x-\mu}{\sigma}$.

The curve for the normal distribution is given by:

$$\phi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

9.3 Joint Distribution

To study the relationship between two or more events, for example:

$$P(A \cap B) = P(X = x \cap Y = y) = P(X = x, Y = y)$$

we can use the **joint probability distribution** to map the intersection. The solutions for random variables X & Y depend on their nature.

9.3.1 Discrete Random Variables

Use the **probability mass function**:

$$f(x, y) = P(X = x, Y = y)$$

If X & Y are independent: $f(x, y) = f(x) \times f(y)$. Whereas if they are dependent, we can use combinations. For example, in a trial with three options (i.e. colour balls in a bag) say, a, b, c, total size t and sample size n we can write:

$$f(x, y) = \frac{\binom{a}{x}\binom{b}{y}\frac{c}{n-x-y}}{{}^{n}C_{n}}$$

To get the marginal distributions:

$$g(x) = \sum_{y} f(x, y) \qquad h(y) = \sum_{x} f(x, y)$$

The expected value of one variable is the sum of each probability in its marginal distribution multiplied by the x or y value, so $E(X) = x_1p_1 + x_2p_2 + \ldots + x_np_n$. The variance is $Var(X) = \sum_x x^2p(x) - [E(X)]^2$.

To measure the relationship between the two variables, we assess the **covariance** and **correlation**:

$$Cov(X, Y) = \sum_{x}\sum_{y} xyf(x, y) - E(X)E(Y)$$

$$\rho_{XY} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

For a discrete pair of random variables X and Y with a joint probability distribution f(x,y), the expected value can be found by use of an arbitrary function of the random variables g(X,Y) such that:

$$E[g(X,Y)] = \sum_x \sum_y g(x,y)\ f(x,y)$$

9.3.2 Continuous Random Variables

Use the **probability density function**:

$$P[(X,Y) \in A] = \int \int_A f(x,y)\ dx\ dy$$

The marginal probability density functions (pdf) are defined by:

$$g(x) = \int_{-\infty}^{\infty} f(x,y)\ dy$$
$$h(y) = \int_{-\infty}^{\infty} f(x,y)\ dx$$

For a continuous pair of random variables the expected value can also be found by use of an arbitrary function:

$$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)\ f(x,y)\ dx\ dy$$

9.3.3 Conditional Distribution

Based on the definition of conditional probability we can draw a conditional distribution based on the joint distribution $f(x,y)$ and the marginal distributions $h(y)$ and $g(x)$ such that:

$$f(x|y) = \frac{f(x,y)}{h(y)}$$

10 Notation

e	base of natural logs (euler's number)
i	square root of -1
∞	infinity
$\forall$	for all
$\exists$	there exists
∇	nabla (del - vector calculus)
$\emptyset$	empty set
$\sum$	sum
$\in$	is an element of / is in
$\subseteq$	is a subset of
$\subset$	is a proper subset of ($\neq$)
$\cap$	intersection
$\cup$	union
$\Rightarrow$	implies that
$\iff$	equivalent to
$\perp$	perpendicular to
$=$	equal to (practical use)
$\equiv$	equivalent / conguent to (theoretical use)

11 Glossary

commutative	an operation is such if altering the order of its operands does not affect its result (i.e. $a + b = b + a$).
distributive	a multiplication is such if $a(b + c) = ab + ac$.
associative	the operation can be regrouped $\equiv (a + b) + c = a + (b + c)$.