Probability and Random Variables; and Classical Estimation Theory
UDRC Summer School, 27th June 2016

Accompanying Notes

Dr James R. Hopgood

Room 2.05
Alexander Graham Bell Building
The King's Buildings
Mayfield Road
Edinburgh
EH9 3JL
Scotland, UK
James.Hopgood@ed.ac.uk
Telephone: +44 (0)131 650 5571
Fax: +44 (0)131 650 6554.

Major revision, Thursday 23rd June, 2016.
Last printed revision with minor corrections, 11 October, 2016.

Typeset by the author with the LaTeX 2_ε Documentation System, with $\mathcal{A}\mathcal{M}\mathcal{S}$-LaTeX Extensions, in 12/18 pt Times and Euler fonts.

INSTITUTE FOR DIGITAL COMMUNICATIONS,
School of Engineering,
College of Science and Engineering,
Kings's Buildings,
Edinburgh, EH9 3JL. U.K.

Copyright Statement

This document does not contain copyright material.

The author of this document

1. holds the copyright for all lecture and course materials in this module;
2. holds the copyright for students notes, summaries, or recordings that substantially reflect the lecture content or materials;
3. makes these materials available only for personal use by students studying this module;
4. reserves the right that no part of the notes, tutorials, solutions, or other course materials may be distributed or reproduced for commercial purposes without express written consent from the author; this does not prevent students from sharing notes on an individual basis for personal use.

These lecture notes consist of entirely original work, where all material has been written and typeset by the author. No figures or substantial pieces of text has been reproduced verbatim from other texts.

However, there is some material that has been based on work in a number of previous textbooks, and therefore some sections and paragraphs have strong similarities in structure and wording. These texts have been referenced and include, amongst a number of others, in order of contributions:

- Manolakis D. G., V. K. Ingle, and S. M. Kogon, *Statistical and Adaptive Signal Processing*: *Spectral Estimation, Signal Modeling, Adaptive Filtering and Array Processing*, McGraw Hill, Inc., 2000.

  IDENTIFIERS – *Paperback*, ISBN10: 0070400512, ISBN13: 9780070400511

- Therrien C. W., *Discrete Random Signals and Statistical Signal Processing*, Prentice-Hall, Inc., 1992.

  IDENTIFIERS – *Paperback*, ISBN10: 0130225452, ISBN13: 9780130225450
  Hardback, ISBN10: 0138521123, ISBN13: 9780138521127

- Kay S. M., *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, Inc., 1993.

  IDENTIFIERS – *Hardback*, ISBN10: 0133457117, ISBN13: 9780133457117
  Paperback, ISBN10: 0130422681, ISBN13: 9780130422682

- Papoulis A. and S. Pillai, *Probability, Random Variables, and Stochastic Processes*, Fourth edition, McGraw Hill, Inc., 2002.

IDENTIFIERS – *Paperback*, ISBN10: 0071226613, ISBN13: 9780071226615
Hardback, ISBN10: 0072817259, ISBN13: 9780072817256

- Proakis J. G. and D. G. Manolakis, *Digital Signal Processing*: *Principles, Algorithms, and Applications*, Pearson New International Edition, Fourth edition, Pearson Education, 2013.

IDENTIFIERS – *Paperback*, ISBN10: 1292025735, ISBN13: 9781292025735

- Mulgew B., P. M. Grant, and J. S. Thompson, *Digital Signal Processing*: *Concepts and Applications*, Palgrave, Macmillan, 2003.

IDENTIFIERS – *Paperback*, ISBN10: 0333963563, ISBN13: 9780333963562

See `http://www.see.ed.ac.uk/~{}pmg/SIGPRO`

- Therrien C. W. and M. Tummala, *Probability and Random Processes for Electrical and Computer Engineers*, Second edition, CRC Press, 2011.

IDENTIFIERS – *Hardback*, ISBN10: 1439826986, ISBN13: 978-1439826980

- Press W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Receipes in C*: *The Art of Scientific Computing*, Second edition, Cambridge University Press, 1992.

IDENTIFIERS – *Paperback*, ISBN10: 0521437202, ISBN13: 9780521437202
Hardback, ISBN10: 0521431085, ISBN13: 9780521431088

The material in [Kay:1993] is mainly covered in Handout 5; material in [Therrien:1992] and [Papoulis:1991] is covered throughout the course. The following labelling convention is used for numbering equations that are taken from the various recommended texts. Equations labelled as:

M:v.w.xyz are similar to those in [Manolakis:2001] with the corresponding label;

T:w.xyz are similar to those in [Therrien:1992] with the corresponding label;

K:w.xyz are similar to those in [Kay:1993] with the corresponding label;

P:v.w.xyz are used in chapters referring to basic digital signal processing (DSP), and are references made to [Proakis:1996].

Contents

Acronyms

2-D	two-dimensional
3-D	three-dimensional
AED	adaptive eigenvalue decomposition
AIC	Akaike's information criterion
AIR	acoustic impulse response
ASL	acoustic source localisation
AVS	acoustic vector sensor
AWGN	additive white Gaussian noise
BFGS	Broyden-Fletcher-Goldfarb-Shannon
BIC	B-Information criterion
BSS	blind source separation
CAT	Parzen's criterion autoregressive transfer function
CCTV	closed-circuit television
CD	compact disc
CLT	central limit theorem
CPSD	cross-power spectral density
CRLB	Cramér-Rao lower-bound
DAT	digital audio tape
DC	"direct current"
DNA	deoxyribonucleic acid
DSP	digital signal processing
DUET	degenerate unmixing estimation technique
DVD	digital versitile disc
DVD-A	digital versitile disc-audio
EEG	electroencephalogram

FPE	final prediction error
FS	Fourier series
FT	Fourier transform
GCC	generalised cross correlation
GCC-PHAT	GCC-phase transform (PHAT)
ICA	independent component analysis
LHS	left hand side
LI	linear intersection
LITP	linear in the parameters
LS	least-squares
LSE	least-squares estimate
LSE	least squares error
LTI	linear time-invariant
MAP	maximum *a posteriori*
MDL	minimum description length
MGF	moment generating function
ML	maximum-likelihood
MLE	maximum-likelihood estimate
MMAP	maximum marginal *a posteriori*
MMSE	minimum mean-square error
MRI	magnetic resonance imaging
MS	mean-square
MSE	mean-squared error
MVU	minimum variance unbiased
MVUE	minimum variance unbiased estimator
NMRI	nuclear magnetic resonance imaging
PHAT	phase transform
PHD	Ph.D. thesis
PTL	passive target localisation
RHS	right hand side

RIR	room impulse response
SACD	super-audio CD
SBF	steered beamformer
SCOT	Smoothed Coherence Transform
SI	spherical interpolation
SRC	stochastic region contraction
SRP	steered response power
SSP	statistical signal processing
STFT	short-time Fourier transform
SX	spherical intersection
TDOA	time-difference of arrival
TF	time-frequency
TFR	time-frequency representation
ULA	uniform linear array
WDO	W-disjoint orthogonality
WGN	white Gaussian noise
cdf	cumulative distribution function
iff	if, and only if,
i. i. d.	independent and identically distributed
i. t. o.	in terms of
pdf	probability density function
pmf	probability mass function
RV	random variable
w. r. t.	with respect to

Acronyms

PET Probability, Random Variables, and Estimation Theory

SSP Statistical Signal Processing

10;

1

Module Overview, Aims and Objectives

Everything that needs to be said has already been said. But since no one was listening, everything must be said again.

André Gide

If you can't explain it simply, you don't understand it well enough.

Albert Einsten

This handout also provides an introduction to signals and systems, and an overview of statistical signal processing applications.

1.1 Obtaining the Latest Version of these Handouts

New slide

- This research tutorial is intended to cover a wide range of aspects which cover the fundamentals of statistical signal processing. It is written at a level which assumes knowledge of undergraduate mathematics and signal processing nomenclature, but otherwise should be accessible to most technical graduates.

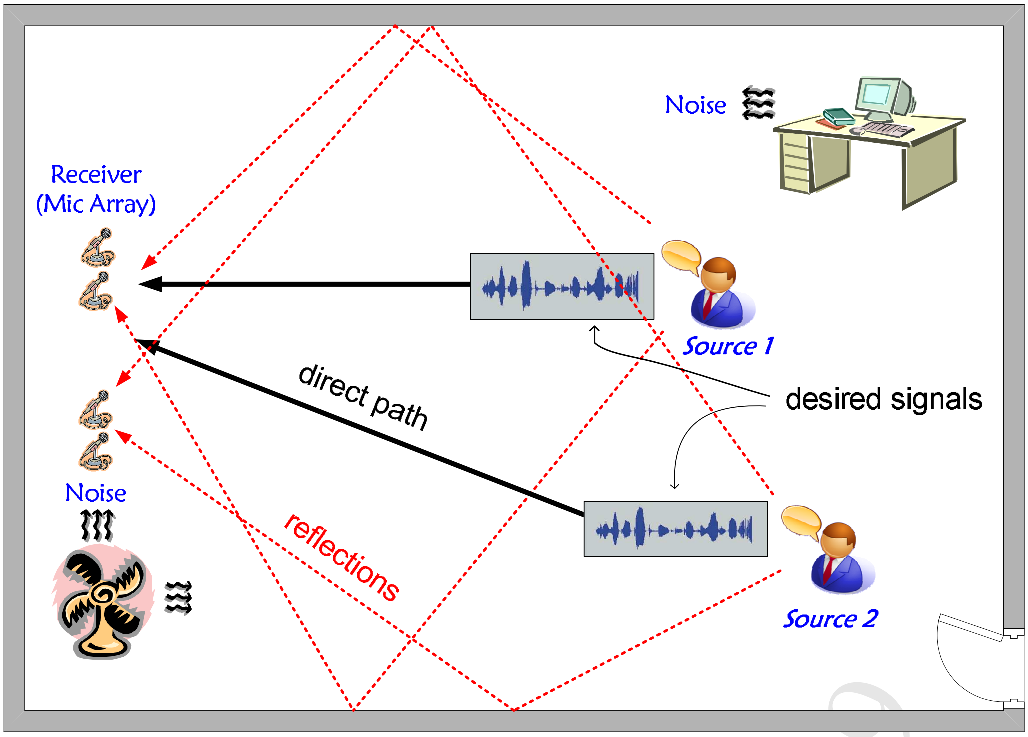


Figure 1.1: Source localisation and BSS. An example of topics using statistical signal processing.

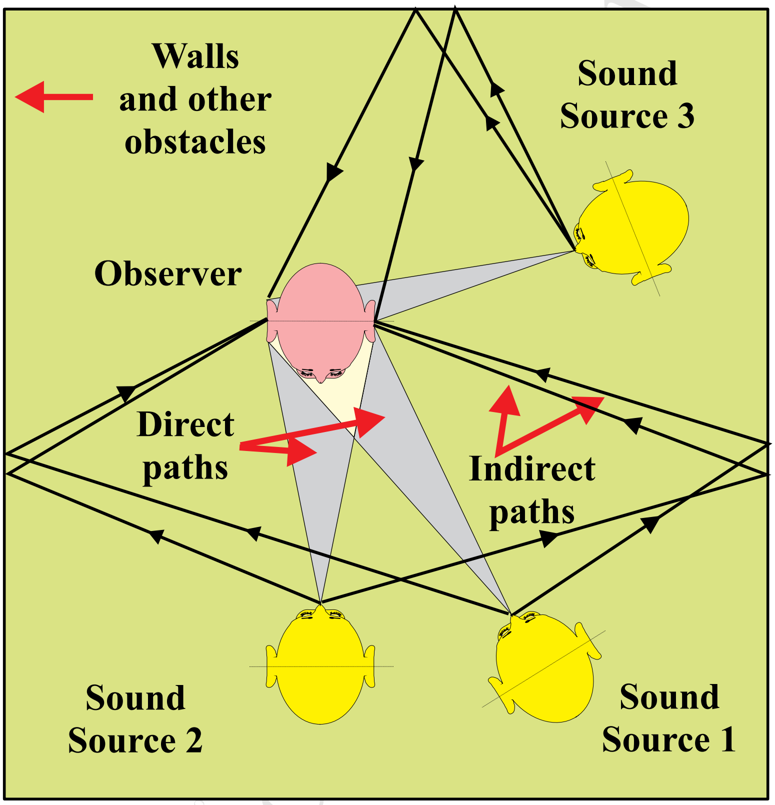


Figure 1.2: Humans turn their head in the direction of interest in order to reduce inteference from other directions; *joint detection, localisation, and enhancement*. An application of probability and estimation theory, and statistical signal processing.

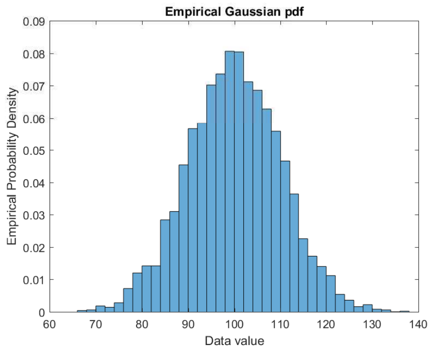


Figure 1.3: Empirical Gaussian probability density function.

KEYPOINT! (Latest Slides). Please note the following:

- This tutorial is being continually updated, and feedback is welcomed. The documents published on the USB stick may differ to the slides presented on the day. In particular, there are likely to be a few typos in the document, so if there is something that isn't clear, please feel free to email me so I can correct it (or make it clearer).

- The latest version of this document can be obtained from the author, Dr James R. Hopgood, by emailing him at: at:

  `mailto:james.hopgood@ed.ac.uk`

  (Update: The notes are no longer online due to the desire to maintain copyright control on the document.)

- Extended thanks are given to the many MSc students over the past 12 years who have helped proof-read and improve these documents.

New slide

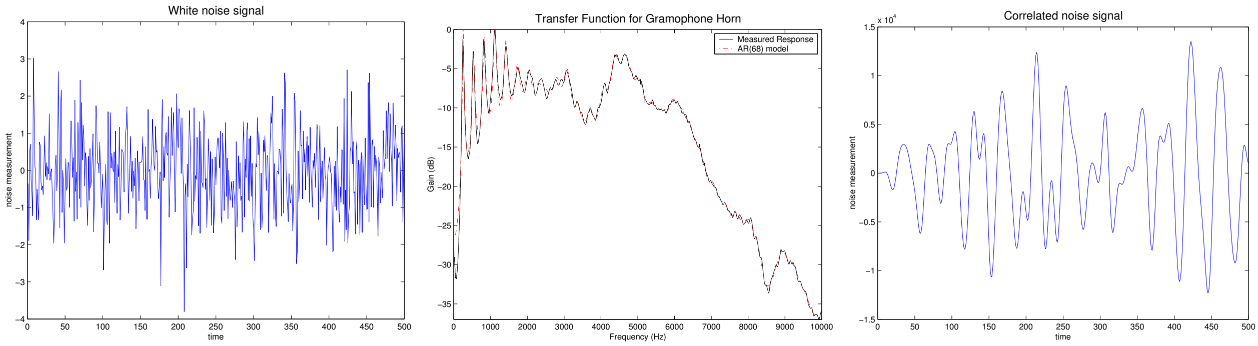
1.2 Module Abstract

The notion of **random** or **stochastic** quantities is an extremely powerful concept that can be constructively used to model observations that result from real-world processes. These quantities could be scalar measurements, such as an instantaneous measurement of distance, or they could be vector-measurements such as a coordinate. They could be random signals either in one-dimension, or in higher-dimensions, such as images. Stochastic quantities such as random signals, by their very nature, are described using the mathematics of probability and statistics. By making assumptions such as the availability of an infinite number of observations or data samples, time-invariant statistics, and known signal or observation models, it is possible to estimate the properties of these random quantities or signals and, consequently, use them in *signal processing* algorithms.

In practice, of course, these statistical properties must be estimated from finite-length data signals observed in noise. In order to understand both the concept of stochastic processes and the inherent **uncertainty** of signal estimates from finite-length sequences, it is first necessary to understand the fundamentals of **probability**, **random variables**, and **estimation theory**.

This topic is covered in two related lecture modules:

1. *Probability, Random Variables, and Estimation Theory*, and



(a) Input signal; uncorrelated white noise process.

(b) Frequency response of channel; the response of an acoustic gramophone horn.

(c) Output signal: a coloured (correlated) noise process.

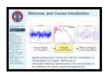
(d) Block diagram of system representing convolution.

Figure 1.4: Solutions to the so-called *blind deconvolution problem* require statistical signal processing methods.

2. *Statistical Signal Processing*,

together introduce the subject of statistical signal modelling and estimation. In particular, the module *Statistical Signal Processing* investigates which statistical properties are relevant for dealing with signal processing problems, how these properties can be estimated from real-world signals, and how they can be used in signal processing algorithms to achieve a particular goal.

1.3 Introduction and Overview

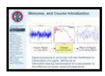


New slide

> Signal processing is concerned with the modification or manipulation of a signal, defined as an information-bearing representation of a real process, to the fulfillment of human needs and aspirations.

Gone is the era where information in the form of electrical signals are processed through analogue devices. For the foreseeable future, processing of digital, sampled, or discrete-time signals is the definitive approach to analysing data and extracting information. In this course, it is assumed that the reader already has a grounding in digital signal processing (DSP), and this module will take you to the next level; a tour of the exciting, fascinating, and active research area of *statistical signal processing (SSP)*.

1.3.1 Description and Learning Outcomes



New slide

Module Aims The aims of the two modules *Probability, Random Variables, and Estimation Theory (PET)*, and *statistical signal processing (SSP)*, are similar to those of the text book [Manolakis:2000, page xvii]. The principle aim of the modules are:

> to provide a unified introduction to the **theory**, **implementation**, and **applications** of statistical signal processing.

Pre-requisites It is strongly recommended that the student has previously attended an undergraduate level course in either signals and systems, digital signal processing, automatic control, or an equivalent course.

Section 1.3.2 provides further details regarding the material a student should have previously covered.

Short Description The **Probability, Random Variables, and Estimation Theory** module introduces the fundamental statistical tools that are required to analyse and describe advanced signal processing algorithms. It provides a unified mathematical framework which is the basis for describing random events and signals, and how to describe key characteristics of random processes.

The module covers probability theory, considers the notion of random variables and vectors, how they can be manipulated, and provides an introduction to estimation theory. It is demonstrated that many estimation problems, and therefore signal processing problems, can be reduced to an exercise in either *optimisation* or *integration*. While these problems can be solved using deterministic numerical methods, the module introduces **Monte Carlo** techniques which are the basis of powerfull stochastic optimisation and integration algorithms. These methods rely on being able to sample numbers, or variates, from arbitrary distributions. This module will therefore discuss the various techniques which are necessary to understand these methods and, if time permits, techniques for random number generation are considered.

The **Statistical Signal Processing** module then consider representing real-world signals by stochastic or random processes. The tools for analysing these random signals are developed in the **Probability, Random Variables, and Estimation Theory** module, and this module extends them to deal with time series. The notion of statistical quantities such as autocorrelation and auto-covariance are extended from random vectors to random processes, and a frequency-domain analysis framework is developed. This module also investigates the affect of systems and transformations on time-series, and how they can be used to help design powerful signal processing algorithms to achieve a particular task.

The module introduces the notion of representing signals using parametric models; it extends the broad topic of statistical estimation theory covered in the **Probability, Random Variables, and Estimation Theory** module for determining optimal model parameters. In particular, the **Bayesian paradigm** for statistical parameter estimation is introduced. Emphasis is placed on relating these concepts to state-of-the-art applications and signals.

Keywords Probability, scalar and multiple random variables, stochastic processes, power spectral densities, linear systems theory, linear signal models, estimation theory, and Monte Carlo methods.

Module Objectives At the end of these modules, a student should be able to have:

1. acquired sufficient expertise in this area to understand and implement **spectral estimation**, **signal modelling**, **parameter estimation**, and **adaptive filtering** techniques;
2. developed an understanding of the basic concepts and methodologies in statistical signal processing that provides the foundation for **further study**, **research**, and **application** to **new problems**.

Intended Learning Outcomes At the end of the **Probability, Random Variables, and Estimation Theory** module, a student should be able to:

1. define, understand and manipulate scalar and multiple random variables, using the theory of probability; this should include the tools of probability transformations and characteristic functions;
2. explain the notion of characterising random variables and random vectors using moments, and be able to manipulate them; understand the relationship between random variables within a random vector;
3. understand the central limit theorem (CLT) and explain its use in estimation theory and the sum of random variables;
4. understand the principles of estimation theory; understand and be apply to apply estimation techniques such as maximum-likelihood, least squares, and Bayesian estimation;
5. be able to characterise the uncertainty in an estimator, as well as characterise the performance of an estimator (bias, variance, and so forth); understand the Cramér-Rao lower-bound (CRLB) and minimum variance unbiased estimator (MVUE) estimators.
6. if time permits, explain and apply methods for generating random numbers, or random variates, from an arbitrary distribution, using methods such as accept-reject and Gibbs sampling; understand the notion of stochastic numerical methods for solving *integration* and *optimisation* problems.

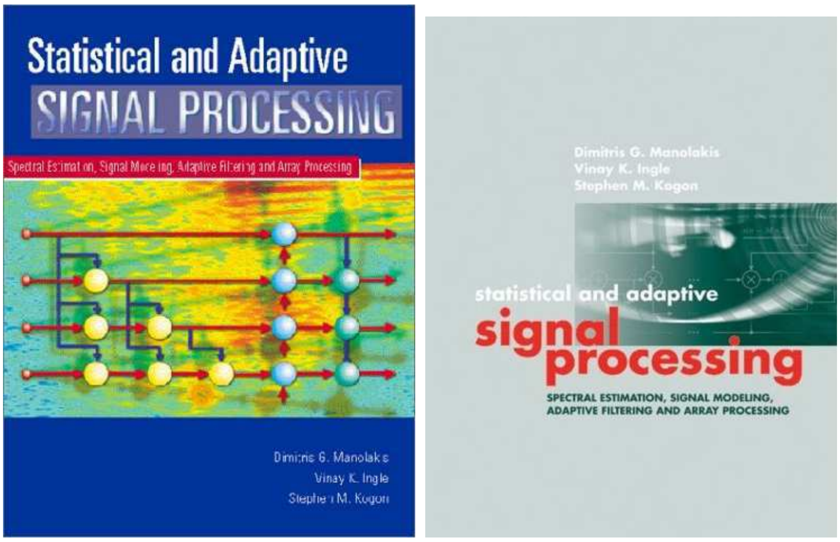
At the end of the **Statistical Signal Processing** module, a student should be able to:

1. explain, describe, and understand the notion of a random process and statistical time series;
2. characterise random processes in terms of its statistical properties, including the notion of stationarity and ergodicity;
3. define, describe, and understand the notion of the power spectral density of stationary random processes; analyse and manipulate power spectral densities;
4. analyse in both time and frequency the affect of transformations and linear systems on random processes, both in terms of the density functions, and statistical moments;
5. explain the notion of parametric signal models, and describe common regression-based signal models in terms of its statistical characteristics, and in terms of its affect on random signals;
6. apply least squares, maximum-likelihood, and Bayesian estimators to model based signal processing problems;

1.3.2 Prerequisites

The mathematical treatment throughout this module is kept at a level that is within the grasp of final-year undergraduate and graduate students, with a background in **digital signal processing (DSP)**, **linear system and control** theory, basic **probability theory**, **calculus**, **linear algebra**, and a competence in Engineering mathematics.

In summary, it is assumed that the reader has knowledge of:



(a) Cover of *paperback* version. (b) Cover of *hardback* version.

Figure 1.5: The main **course text** for this module: [Manolakis:2000].

1. Engineering mathematics, including linear algebra, manipulation of vectors and matrices, complex numbers, linear transforms including Fourier series and Fourier transforms, z-transforms, and Laplace transforms;

2. basic probability and statistics, albeit with a solid understanding;

3. differential and integral calculus, including differentiating products and quotients, functions of functions, integration by parts, integration by substitution;

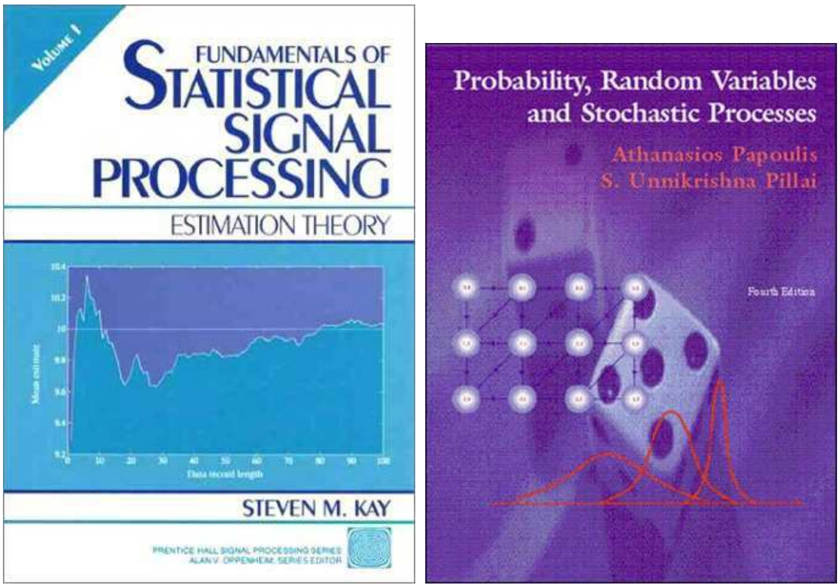
4. basic digital signal processing (DSP), including:

   - the notions of deterministic continuous-time signals, discrete-time signals and digital (quantised) signals;
   - filtering and inverse filtering of signals; convolution;
   - the response of linear systems to harmonic inputs; analysing the time and frequency domain properties of signals and systems;
   - sampling of continuous time processes, Nyquist's sampling theorem and signal reconstruction;
   - and analysing discrete-time signals and systems.

Note that while the reader should have been exposed to the idea of a **random variable**, it is **not** assumed that the reader has been introduced to *random signals* in any form. A list of recommended texts covering these prerequisites is given in Section 1.3.3.

1.3.3 Recommended Texts for Module Content

The **recommended text** for this module is cited throughout this document as [Manolakis:2000]. The full reference is:



(a) **Recommended text**: [Kay:1993].

(b) **Recommended text**: [Papoulis:1991].

Figure 1.6: Additional recommended texts for the course.

Manolakis D. G., V. K. Ingle, and S. M. Kogon, *Statistical and Adaptive Signal Processing*: *Spectral Estimation, Signal Modeling, Adaptive Filtering and Array Processing*, McGraw Hill, Inc., 2000.

IDENTIFIERS – *Paperback*, ISBN10: 0070400512, ISBN13: 9780070400511

It is recommended that, if you wish to purchase a hard-copy of this book, you try and find this paperback version; it should be possible to order a copy relatively cheaply through the US version of Amazon (check shipping costs). However, please note that this book is now available, at great expense, in hard-back from an alternative publisher. The full reference is:

Manolakis D. G., V. K. Ingle, and S. M. Kogon, *Statistical and Adaptive Signal Processing*: *Spectral Estimation, Signal Modeling, Adaptive Filtering and Array Processing*, Artech House, 2005.

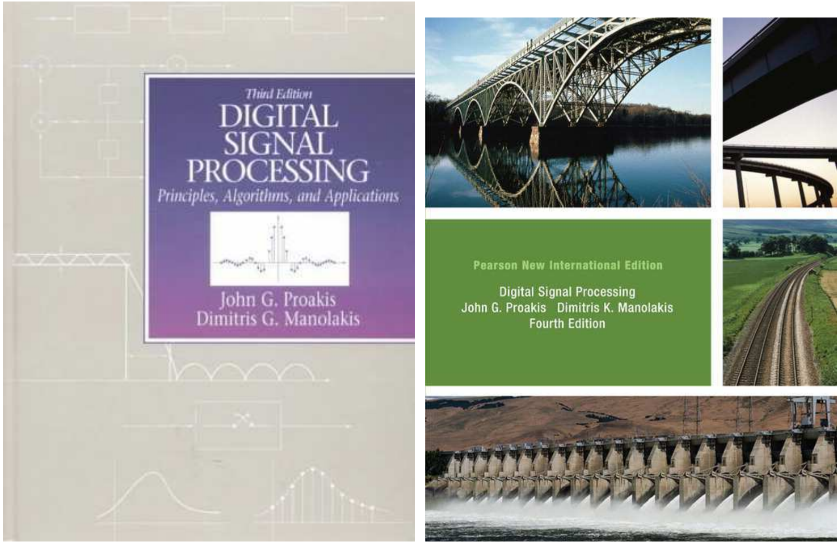
IDENTIFIERS – *Hardback*, ISBN10: 1580536107, ISBN13: 9781580536103

Images of the book covers are shown in Figure 1.5. For further reading, or an alternative perspective on the subject matter, other recommended text books for this module include:

1. Therrien C. W., *Discrete Random Signals and Statistical Signal Processing*, Prentice-Hall, Inc., 1992.

IDENTIFIERS – *Paperback*, ISBN10: 0130225452, ISBN13: 9780130225450
Hardback, ISBN10: 0138521123, ISBN13: 9780138521127

2. Kay S. M., *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, Inc., 1993.



(a) Third Edition cover. (b) Fourth Edition cover.

Figure 1.7: **Course text**: further reading for digital signal processing and mathematics, [Proakis:1996].

IDENTIFIERS – *Hardback*, ISBN10: 0133457117, ISBN13: 9780133457117
Paperback, ISBN10: 0130422681, ISBN13: 9780130422682

3. Papoulis A. and S. Pillai, *Probability, Random Variables, and Stochastic Processes*, Fourth edition, McGraw Hill, Inc., 2002.

IDENTIFIERS – *Paperback*, ISBN10: 0071226613, ISBN13: 9780071226615
Hardback, ISBN10: 0072817259, ISBN13: 9780072817256

These are referenced throughout as [Therrien:1992], [Kay:1993], and [Papoulis:1991], respectively. Images of the book covers are shown in Figure 1.6. The material in [Kay:1993] is mainly covered in Handout 5 on Estimation Theory of the PET module. The material in [Therrien:1992] and [Papoulis:1991] is covered throughout the course, with the former primarily in the Statistical Signal Processing (SSP) module.

KEYPOINT! (Proposed Recommended Text Book for Future Years). Finally, Therrien has also published a recent book which covers much of this course extremely well, and therefore comes thoroughly recommended. It has a number of excellent examples, and covers the material in good detail.

Therrien C. W. and M. Tummala, *Probability and Random Processes for Electrical and Computer Engineers*, Second edition, CRC Press, 2011.

IDENTIFIERS – *Hardback*, ISBN10: 1439826986, ISBN13: 978-1439826980

1.3.4 Recommended Texts: Prerequisite Material

As mentioned in Section 1.3.2 above, regarding the prerequisites, it is assumed that the reader has a basic knowledge of digital signal processing. If not, or if the reader wishes to revise the topic, the following book which is *highly* recommended:

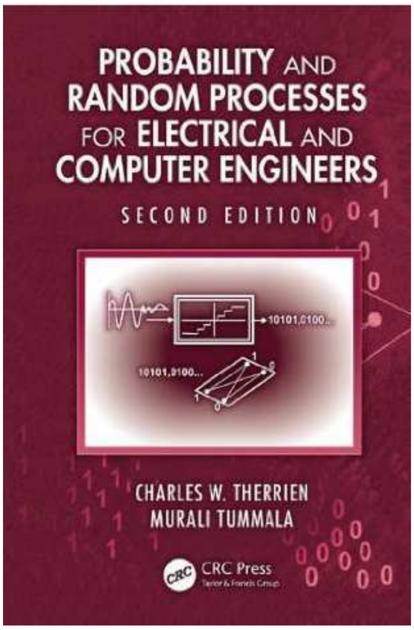


Figure 1.8: Further reading for statistical signal processing, [Therrien:2011].

Proakis J. G. and D. G. Manolakis, *Digital Signal Processing*: *Principles, Algorithms, and Applications*, Third edition, Prentice-Hall, Inc., 1996.

IDENTIFIERS – *Paperback*, ISBN10: 0133942899, ISBN13: 9780133942897
Hardback, ISBN10: 0133737624, ISBN13: 9780133737622

This is cited throughout as [Proakis:1996] and is referred to in the second handout. This is the *third edition* to the book, and a *fourth edition has recently been released*:

Proakis J. G. and D. G. Manolakis, *Digital Signal Processing*: *Principles, Algorithms, and Applications*, Pearson New International Edition, Fourth edition, Pearson Education, 2013.

IDENTIFIERS – *Paperback*, ISBN10: 1292025735, ISBN13: 9781292025735

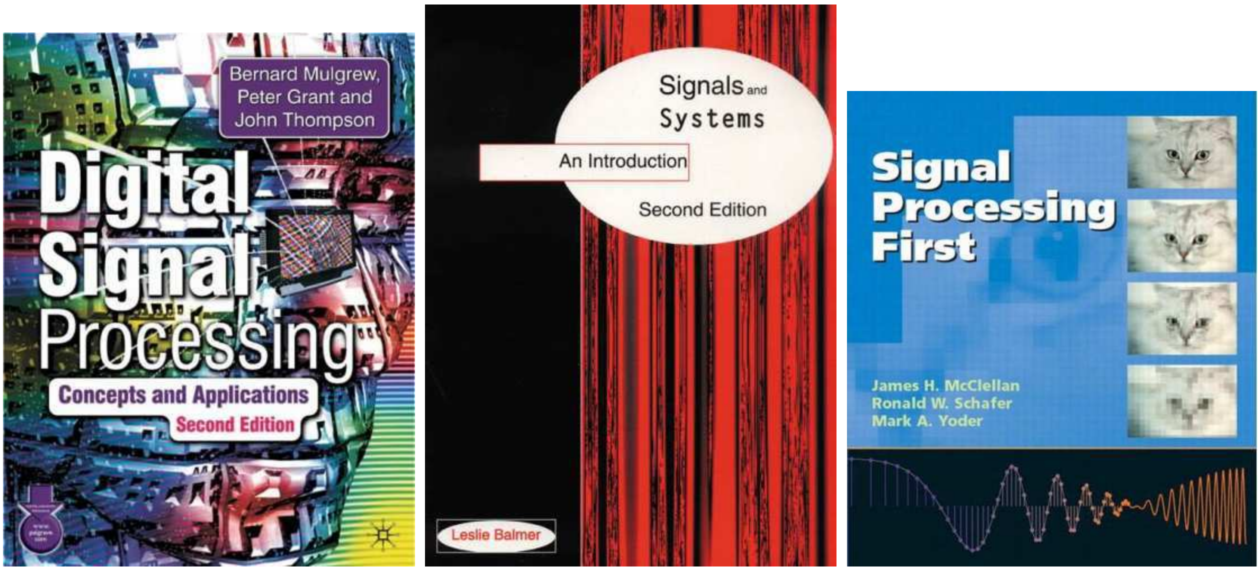
Although it is best to purchase the *fourth edition*, please bear in mind that the equation references throughout the lecture notes correspond to the third edition. For an undergraduate level text book covering an introduction to signals and systems theory, which it is assumed you have covered, the following is recommended [Mulgrew:2002]:

Mulgew B., P. M. Grant, and J. S. Thompson, *Digital Signal Processing*: *Concepts and Applications*, Palgrave, Macmillan, 2003.

IDENTIFIERS – *Paperback*, ISBN10: 0333963563, ISBN13: 9780333963562

See `http://www.see.ed.ac.uk/~{}pmg/SIGPRO`

The latest edition was printed in 2003, but any of the book edition will do. An alternative presentation of roughly the same material is provided by the following book [Balmer:1997]:



(a) [Mulgrew:2002]. (b) [Balmer:1997]. (c) [McClennan:2003].

Figure 1.9: Undergraduate texts on Signals and Systems.

Balmer L., *Signals and Systems*: *An Introduction*, Second edition, Prentice-Hall, Inc., 1997.

IDENTIFIERS – *Paperback*, ISBN10: 0134954729, ISBN13: 9780134956725

The Appendix on complex numbers may prove useful.

For an excellent and gentle introduction to signals and systems, with an elegant yet thorough overview of the mathematical framework involved, have a look at the following book, if you can get hold of a copy (but don't go spending money on it):

McClellan J. H., R. W. Schafer, and M. A. Yoder, *Signal Processing First*, Pearson Education, Inv, 2003.

IDENTIFIERS – *Paperback*, ISBN10: 0131202650, ISBN13: 9780131202658
Hardback, ISBN10: 0130909998, ISBN13: 9780130909992

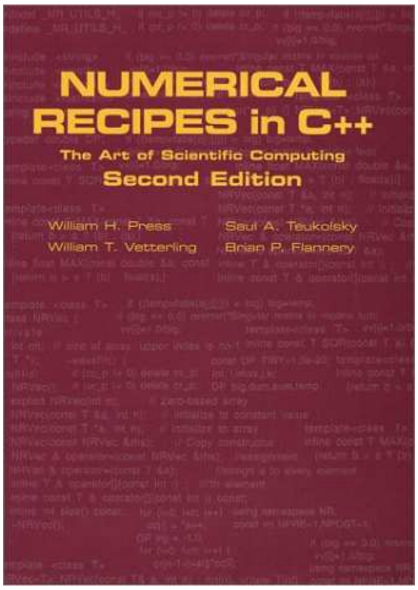
1.3.5 Further Recommended Reading

For additional reading, and for guides to the implementation of numerical algorithms used for some of the actual calculations in this lecture course, the following book is also strongly recommended:

Press W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Receipes in C*: *The Art of Scientific Computing*, Second edition, Cambridge University Press, 1992.

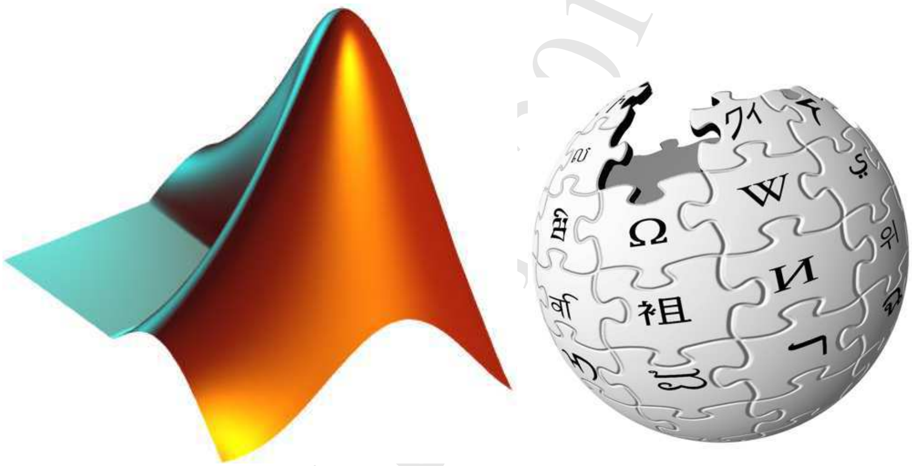
IDENTIFIERS – *Paperback*, ISBN10: 0521437202, ISBN13: 9780521437202
Hardback, ISBN10: 0521431085, ISBN13: 9780521431088

Please note that there are many versions of the *numerical recipes* book, and that any version will do. So it would be worth getting the latest version.



(a) **Recommended text**: [Press:1992].

Figure 1.10: Further reading for numerical methods and mathematics.



(a) The MATLAB logo. MATLAB is a useful utility to experiment with.

(b) Wikipedia, The Free Encyclopedia.

Figure 1.11: Some useful resources.

1.3.6 Additional Resources

Other useful resources include:

- The extremely comprehensive and interactive mathematics encyclopedia:

  Weisstein E. W., *MathWorld*, From MathWorld - A Wolfram Web Resource, 2008.

  See `http://mathworld.wolfram.com`

- Connexions is an environment for collaboratively developing, freely sharing, and rapidly publishing scholarly content on the Web. A wide variety of technical lectures can be found at:

  Connexions, The Connexions Project, 2008.

  See `http://cnx.org`

- The Wikipedia online encyclopedia is very useful, although beware that there is no guarantee that the technical articles are either correct, or comprehensive. However, there are some excellent articles available on the site, so it is worth taking a look.

  *Wikipedia, The Free Encyclopedia*Wikipedia, The Free Encyclopedia, 2001 – present.

  See `http://en.wikipedia.org/`

- The Mathworks website, the creators of MATLAB, contains much useful information:

  MATLAB: The language of technical computing, The MathWorks, Inc., 2008.

  See `http://www.mathworks.com/`

- And, of course, the one website to rule them all:

  Google Search Engine, Google, Inc., 1998 – present.

  See `http://www.google.co.uk`

1.3.7 Convention for Equation Numbering

In this handout, the following labelling convention is used for numbering equations that are taken from the various recommended texts. This labelling should be helpful for locating the relevant sections in the books for further reading. Equations labelled as:

M:v.w.xyz are similar to those with the same equation reference in the core recommended text book, namely [Manolakis:2001];

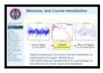
T:w.xyz are similar to those in [Therrien:1992] with the corresponding label;

K:w.xyz are similar to those in [Kay:1993] with the corresponding label;

P:v.w.xyz are used in chapters referring to basic DSP, and are references made to [Proakis:1996].

All other equation labeling refers to intra-cross-referencing for these handouts. Most equations are numbered for ease of referencing the equations, should you wish to refer to them in tutorials or email communications, and so forth.

1.4 What are Signals and Systems?



New slide

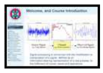
Common usage and understanding of the word *signal* is actually correct from an Engineering perspective within some rather broad definitions: a signal is thought of as *something* that carries information. Usually, that *something* is a pattern of variations of a physical quantity that can be manipulated, stored, or transmitted by a physical process. Examples include speech signals, general audio signals, video or image signals, biomedical signals, radar signals, and seismic signals, to name but a few.

So formally, a **signal** is defined as an information-bearing representation of a real physical process. It is important to recognise that signals can take many equivalent forms or representations. For example, a speech signal is produced as an acoustic signal, but it can be converted to an electrical signal by a microphone, or a pattern of magnetization on a magnetic tape, or even as a string of numbers as in digital audio recording.

The term *system* is a little more ambiguous, and can be subject to interpretation. The word *system* can correctly be understood as a process, but often the word *system* is used to refer to a large organisation that administers or implements some process.

In Engineering terminology, a **system** is something that can manipulate, change, record, or transmit **signal**s. In general, **system**s operate on **signal**s to produce new signals or new signal representations. For example, an audio compact disc (CD) stores or represents a music signal as a sequence of numbers. A CD player is a system for converting the numerical representation of the signal stored on the disk to an acoustic signal that can be heard.

1.4.1 Mathematical Representation of Signals

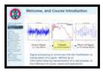


New slide

A *signal* is defined as an information-bearing representation of a real process. It is a pattern of variations, commonly referred to as a waveform, that encodes, represents, and carries information.

Many signals are naturally thought of as a pattern of variations with time. For example, a speech signal arises as a pattern of changing air pressure in the vocal tract, creating a sound wave, which is then converted into electrical energy using a microphone. This electrical signal can then be plotted as a time-waveform, and an example is shown in Figure 1.12. The vertical axis denotes air pressure or microphone voltage, and the horizontal axis represents time. This particular plot shows four contiguous segments of the speech waveform. The second plot is a continuation of the first, and so on, and each plot is vertically offset with the starting time of each segment shown on the left vertical axis.

1.4.1.1 Continuous-time and discrete-time signals



New slide

The signal shown in Figure 1.12 is an example of a one-dimensional **continuous-time signal**. Such signals can be represented mathematically as a function of a single independent variable, t, which represents time and can take on any real-valued number. Hence, each segment of the speech waveform can be associated with a function $s(t)$. In some cases, the function $s(t)$ might be a simple function, such as a sinusoid, but for real signals, it will be a complicated function.

Generally, most *real world* signals are continuous in time and analogue: this means they exist for all time-instances, and can assume any value, within a predefined range, at these time instances. Although most signals originate as continuous-time signals, digital processors and devices can only deal with **discrete-time signals**. A discrete-time representation of a signal can be obtained from a continuous-time signal by a process known as **sampling**. There is an elegant theoretical foundation

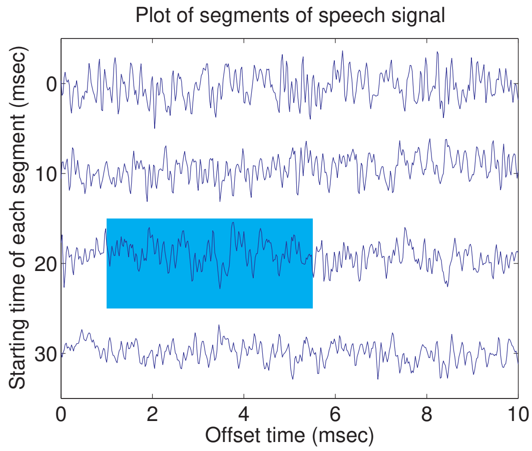


Figure 1.12: Plot of part of a speech signal. This signal can be represented by the function $s(t)$, where t is the independent variable representing time. The shaded region is shown in more detail in Figure 1.13.

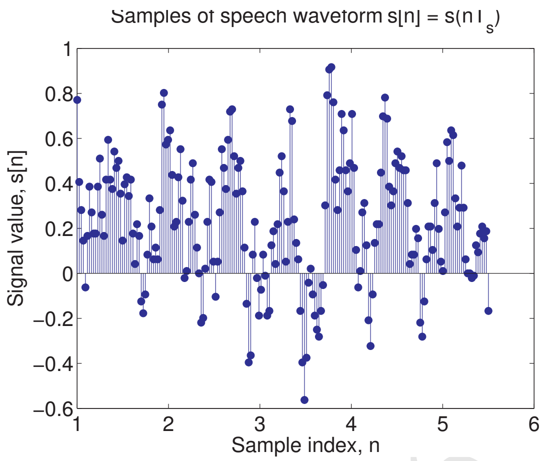


Figure 1.13: Example of a discrete-time signal. This is a sampled version of the shaded region shown in Figure 1.12.

to the process of sampling, although it suffices to say that the result of sampling a continuous-time signal at isolated, equally spaced points in time is a sequence of numbers that can be represented as a function of an index variable that can take on only discrete integer values.

The sampling points are spaced by the **sampling period**, denoted by T_s. Hence, the continuous-time signal, $s(t)$, is *sampled* at times $t = nT_s$ resulting in the discrete-time waveform denoted by:

$$s[n] = s(nT_s), \quad n \in \{0, 1, 2, \dots\}. \tag{1.1}$$

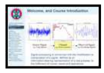
where n is the index variable. A discrete-time signal is sometimes referred to as a discrete-time sequence, since the waveform $s[n]$ is a sequence of numbers. Note, the convention that parenthesis $(\,)$ are used to enclose the independent variable of a continuous-time function, and square brackets $[\,]$ enclose the index variable of a discrete-time signal. Unfortunately, this notation is not always adhered too (and is not yet consistent in these notes either).

Figure 1.13 shows an example of a short segment of the speech waveform from Figure 1.12, with a sampling period of $T_s = \frac{1}{44100}$ seconds, or a sampling frequency of $f_s = \frac{1}{T_s} = 44.1$ kHz. It is not possible to evaluate the continuous-time function $s(t)$ for every value of t, only at a finite-set of points, which will take a finite time to evaluate. Intuitively, however, it is known that the closer the

Figure 1.14: Example of a signal that can be represented by a function of two spatial variables.

spacing of the sampled points, the more the sequence retains the shape of the original continuous-time signal. The question arises, then, regarding what is the largest **sampling period** that can be used to retain all or most of the information about the original signal.

1.4.1.2 Other types of signals



New slide

While many signals can be considered as evolving patterns in time, many other signals are not time-varying patterns at all. For example, an image formed by focusing light through a lens onto an imaging array is a spatial pattern. Thus, an image is represented mathematically as a function of two independent spatial variables, x and y; thus, a picture might be denoted as $p(x, y)$. An example of a **gray-scale image** is shown in Figure 1.14; thus, the value $p(x_0, y_0)$ represents the particular shade of gray at position (x_0, y_0) in the image.

Although images such as that shown in Figure 1.14 represents a quantity from a physical two-dimensional (2-D) spatial continuum, digital images are usually discrete-variable 2-D signals obtained by sampling a continuous-variable 2-D signal. Such a 2-D discrete-variable signal would be represented by a 2-D sequence or array of numbers, and is denoted by:

$$p[m,\, n] = p(m\Delta_x,\, n\Delta_y), \quad m, n \in \{0,\, 1,\, \dots\, N-1\}. \tag{1.2}$$

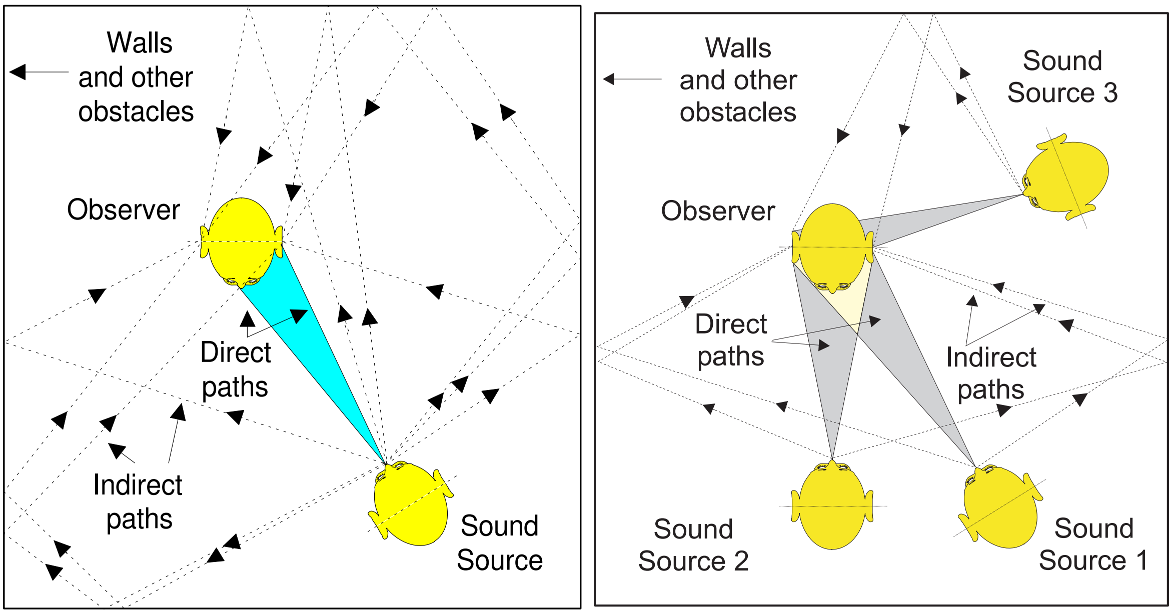
where m and n take on integer values, and Δ_x and Δ_y are the horizontal and vertical sampling spacing or periods, respectively.

Two-dimensional functions are appropriate mathematical representations of still images that do not change with time; on the other hand, a sequence of images that creates a video requires a third independent variable for time. Thus, a video sequence is represented by the three-dimensional (3-D) function $v(x,\, y,\, t)$.

The purpose of this section is to introduce the idea that signals can:

- be represented by mathematical functions in one or more dimensions;
- be functions of continuous or discrete variables.

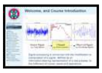
The connection between functions and signals is key to signal processing and, at this point, functions serve as abstract symbols for signals. This is an important, but very simple, concept for using mathematics to describe signals and systems in a systematic way.



(a) Acoustic path from a sound source to a microphone.

(b) Many sound sources within a room.

Figure 1.15: Observed signals in room acoustics.



New slide

1.4.2 Mathematical Representation of Systems

A **system** manipulates, changes, records, or transmits **signal**s. To be more specific, a one-dimensional continuous-time system takes an input signal $x(t)$ and produces a corresponding output signal $y(t)$. This can be represented mathematically by the expression

$$y(t) = \mathcal{T}\{x(t)\} \tag{1.3}$$

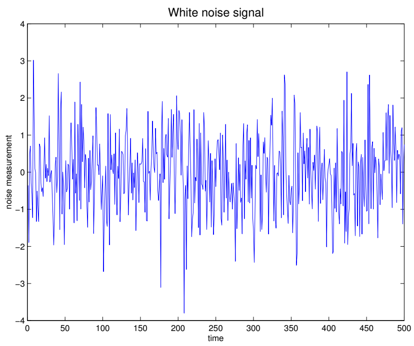
which means that the input signal, $x(t)$, be it a waveform or an image, is operated on by the system, which is symbolised by the operator $\mathcal{T}$ to produce the output $y(t)$. So, for example, consider a signal that is the square of the input signal; this is represented by the equation

$$y(t) = [x(t)]^2 \tag{1.4}$$

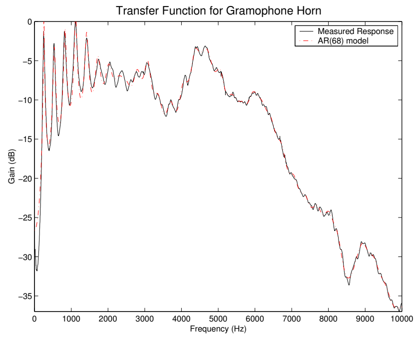
Figure 1.15 and Figure 1.17 show how signals can be generated and observed in a real application. In Figure 1.15, the sound source and the information received by the observer, or microphone, are the **signals**; the room acoustics represent the **system**. Figure 1.16 shows the **input signal** to the system, a *characterisation of the system*, and the resulting **output signal**. In Figure 1.17, the blurred images are the result of the original image being passed through a **linear system**; the linear system represents the physical process of a camera, for example, being out-of-focus, or in motion relative to the object of interest.

The subject of signals and systems is the basis of a branch of Engineering known as signal processing; this area is formally defined as follows:

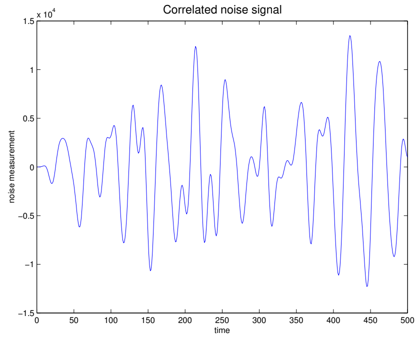
> *Signal processing* is concerned with the modification or manipulation of a signal, defined as an information-bearing representation of a real process, that has been passed through a *system*, to the fulfillment of human needs and aspirations.



(a) Source signal *into* a system.



(b) A frequency response *representing* the characteristics of the system.



(c) The system output.

(d) Block diagram representation of signal paths.

Figure 1.16: The result of passing a signal through a system.

(a) An original unblurred noiseless image.

(b) An image distorted by an out-of-focus blur.

(c) Image distorted by motion blur.

Figure 1.17: A blind image deconvolution problem; restoration of natural photographic images.

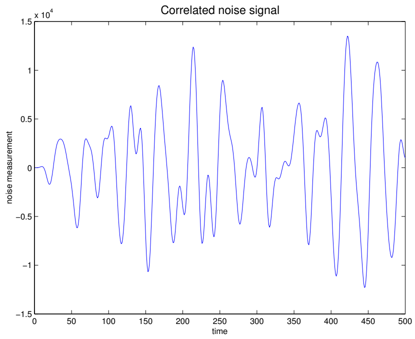
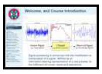


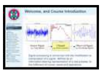
Figure 1.18: Amplitude-verses-time plot.



New slide

1.4.3 Deterministic Signals

The deterministic signal model assumes that signals are explicitly known for all time from time $t = -\infty$ to $t = +\infty$, where $t \in \mathbb{R}$, the set of all real numbers. There is absolutely no uncertainty whatsoever regarding their past, present, or future signal values. The simplest description of such signals is an amplitude-verses-time plot, such as that shown in Figure 1.18; this *time history* helps in the identification of specific patterns, which can subsequently be used to extract information from the signal. However, quite often, information present in a signal becomes more evident by transformation of the signal into another domain, and one of the most nature examples is the frequency domain.



New slide

1.5 Motivation for Signal Modelling

Some state-of-the-art applications of statistical signal processing include the following:

Biomedical From medical imaging to analysis and diagnosis, signal processing is now dominant in patient monitoring, preventive health care, and tele-medicine. From analysing electroencephalogram (EEG) scans to magnetic resonance imaging (MRI) (or nuclear magnetic resonance imaging (NMRI)), to classification and analysis of deoxyribonucleic acid (DNA) from micro-arrays, signal processing is required to make sense of the analogue signals to then provide information to clinicians and doctors.

Surveillance and homeland security From fingerprint analysis, voice transcription and communication monitoring, to the analysis of closed-circuit television (CCTV) footage, digital signal processing is applied in many areas of homeland security. It is an especially well-funded area at the moment.

Target tracking and navigation Although radar and sonar principally use analogue signals for *illuminating* an object with either an electromagnetic or acoustic wave, discrete-time signal processing is the primary method for analysing the received data. Typical features for estimation include detecting targets, estimating the position, orientation, and velocity of the object, target tracking and target identification.

Of recent interest is tracking groups of targets, such as a convoy of vehicles, or a flock of birds. Attempting to track each individual target is an overly complicated problem, and by considering the group dynamics of a particular scenario, the multi-target tracking problem is substantially simplified.

Mobile communications New challenges in mobile communications include next-generation networks; users demand higher data-rates which, in-turn, requires higher bandwidth. Typically, higher-bandwidth communication systems have shorter range. Rather than have more and more base stations for the mobile network, there is substantial research into mobile ad-hoc networks.

A mobile ad-hoc network is a self-configuring network of mobile routers connected by wireless links, forming an arbitrary topology. The routers are free to move randomly and organize themselves arbitrarily; thus, the network's wireless topology may change rapidly and unpredictably. The challenge is to design a system that can cope with this changing topology, and is a very active area of research in communication theory.

A testament to the change in mobile communications is the availability of cheap mobile broadband modems which provide broadband Internet access which is comparable with fixed-line technologies that were available only a few years ago.

Speech enhancement and recognition Whether for the analysis of a black-box recording, for enhancing speech recognition in noisy and reverberant environments, or for the improved acoustic clarity of mobile phone conversations, the enhancement of acoustic signals is still a major aspect of signal processing research.

Many signal processing systems are designed to extract information for some purpose. They share the common problem of needing to estimate the values of a group of parameters. Such algorithms involve signal modelling and spectral estimation. Some typical applications and the desired parameter include:

Radar Radar is primarily used in determining the position of an aircraft or other moving object; for example, in airport surveillance. It is desirable to estimate the range of the aircraft, as determined by the time for an electromagnetic pulse to be reflected by the aircraft.

Sonar Sonar is also interested in the position of a target, such as a submarine. However, whereas radar is, mostly, an *active* device in the sense that it transmits an electromagnetic pulse to *illuminate* the target, sonar listens for noise radiated by the target. This radiated noise includes sounds generated by machinery, or the propeller action. Then, by using a **sensor array** where the relative positions of each sensor are known, the time delay between the arrival of the pulse at each sensor can be measured and this can be used to determine the bearing of the target.

Image analysis It might be desirable to estimate the position and orientation of an object from a camera image. This would be useful, for example, in guiding a robot to pick up an object. Alternatively, it might be desirable to remove various forms of blur from an image, as shown in Figure 1.17; this blur might be characterised by a parametric function.

Biomedicine A parameter of interest might be the heart rate of a fetus.

Communications Estimate the carrier frequency of a signal such that the signal can be demodulated to baseband.

Control Estimate the position of a boat such that corrective navigational action can be taken.

Seismology Estimate the underground distance of an oil deposit based on sound reflections due to different densities of oil and rock layers.

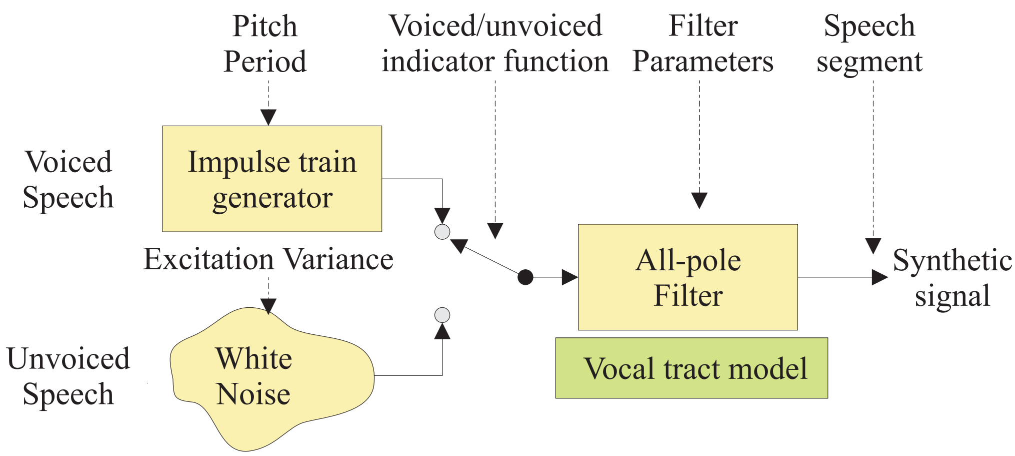
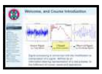


Figure 1.19: The speech synthesis model.

And the list can go on, with a multitude of applications stemming from the analysis of data from physical experiments through to economic analysis. To gain some motivation for looking at various aspects of statistical signal processing, some specific applications will be considered that require the tools this module will introduce. These applications include:

- Speech Modelling and Recognition
- Single Channel Blind System Identification
- Blind Signal Separation
- Data Compression
- Enhancement of Signals in Noise



New slide

1.5.1 Speech Modelling and Recognition

Statistical parametric modelling can be used to characterise the speech production system, and therefore can be applied in the analysis and synthesis of speech. In the analysis of speech, the waveform is sampled at a rate of about 8 to 20 kHz, and broken up into short segments whose duration is typically 10 to 20 ms; this results in consecutive segments containing about 80 to 400 time samples.

Most speech sounds, generally, are classified as either *voiced* or *unvoiced* speech:

- voiced speech is characteristic of vowels;
- unvoiced speech is characteristic of consonants at the beginning of syllables, fricatives (/f/, /s/ sounds), and a combination of these.

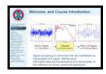
Thinking of the types of sound fields created by vowels, it is apparent that *voiced speech* has a harmonic quality. In fact, it is sometimes known as frequency-modulated speech. A commonly used model for voiced speech exploits this harmonic characteristic, and uses the so-called *sum-of-sinusoids* decomposition. *Unvoiced speech*, on the other hand, does not exhibit such a harmonic structure, although it does possesses a form that can be modelled using the statistical models introduced in later lectures.

Figure 1.20: Solutions to the blind deconvolution problem requires advanced statistical signal processing.

For both of these types of speech, the production is modelled by driving or exciting a linear system, representing the vocal tract, with an excitation having a flat (or constant) spectrum.

The vocal tract, in turn, is modelled by using a pole-zero system, with the poles modelling the vocal tract resonances and the zeros serving the purpose of dampening the spectral response between pole frequencies. In the case of voiced speech, the input to the vocal tract model is a quasi-periodic pulse waveform, whereas for unvoiced speech, the source is modelled as random noise. Thus, the complete set of parameters for this model include an indicator variable as to whether the speech is voiced or unvoiced, the pitch period for voiced sounds, the gain or variance parameter for unvoiced sounds, and the coefficients for the all-pole filter modelling the vocal tract filter. The model is shown in Figure 1.19. This model is widely used for low-bit-rate (less than 2.4 kbits/s) **speech coding**, **synthetic speech generation**, and extraction of features for speaker and **speech recognition**.

1.5.2 Single Channel Blind System Identification



New slide

Consider the following abstract problem that is shown in Figure 1.20:

- The output only of a system is observed, and it is desirable to estimate the source signal that is applied to the input of the system without knowledge of the system itself. In other-words, the output observation, $\mathbf{x} = \{x[n],\, n \in \mathbb{Z}\}$,[1] is modelled as a function of the unknown source signal, $\mathbf{s} = \{s[n],\, n \in \mathbb{Z}\}$, with an unknown, possibly nonlinear, distortion denoted by $\mathcal{F}$; more formally, $\mathbf{x} = \mathcal{F}(\mathbf{s})$.

- When the function $\mathcal{F}$ is linear time-invariant (LTI), and defined by the impulse response $h[n]$, then:

$$x[n] = h[n] * s[n] = \sum_{k \in \mathbb{Z}} h[n-k]\, s[n] \tag{1.5}$$

- **Problem**: Given only $\{x[n]\}$, estimate either the channel function, $\mathcal{F}$, which in the LTI case will be represented by the impulse response $h[n]$, *or* a scaled shifted version of the source signal, $\{s[n]\}$; i.e. $\hat{s}[n] = a\, s[n-l]$ for some l.

The distortion operator, $\mathcal{F}$, could represent the:

- acoustical properties of a room (with applications in **hands free telephones**, **hearing aids**, **archive restoration**, and **automatic speech recognition**);
- effect of multi-path radio propagation (with applications in **communication channels**);
- non-impulsive excitation in seismic applications (with applications in **seismology**);
- blurring functions in **image processing**; in this case, the signals are 2-D.

[1] The notation $n \in \mathbb{Z}$ means that n belongs to, or is an element of, the set of integers: $\{-\infty, \dots, -2, -1, 0, 1, 2, \dots, \infty\}$. In otherwords, it may take on any integer value.

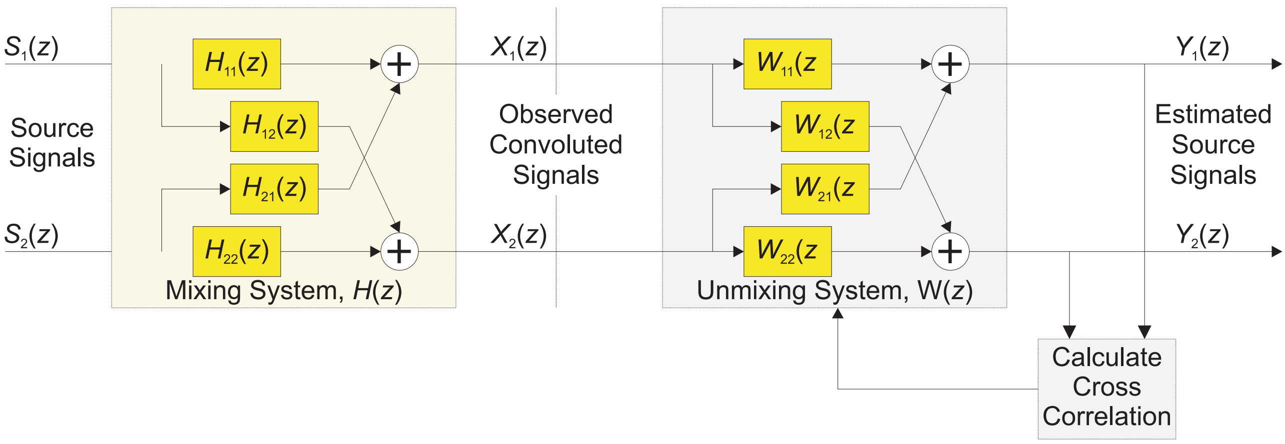
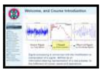


Figure 1.21: Standard signal separation using the independent component assumption.

This problem can only be solved by parametrically modelling the source signal and channel, and using **parameter estimation** techniques to determine the appropriate parameter values.



New slide

1.5.3 Blind Signal Separation

An extremely broad and fundamental problem in signal processing is BSS, and an important special case is the separation of a mixture of audio signals in an acoustic environment. Typical applications include the separation of overlapping speech signals, the separation of musical instruments, enhancement of speech recordings in the presence of background sounds, or any variation of the three. In general, a number of sounds at discrete locations within a room are filtered due to room acoustics and then mixed at the observation points; for example, a microphone will pick up a number of reverberant sounds simultaneously (see Figure 1.15).

A very powerful paradigm within which signal separation can be achieved is the assumption that the source signals are statistically independent of one another; this is known as independent component analysis (ICA). Figure 1.21 demonstrates a separation algorithm based on ICA; an “unmixing” system is chosen that has minimal statistical correlation (a sufficient but not necessary condition for statistical independence, as will be seen later in this course) of the hypothesised separated signals, thereby matching the statistical characteristics of the original signals. This algorithm then uses standard convex optimisation algorithms to solve the minimisation problem.

It is clear, then, that this approach to ICA requires good estimates of the correlation functions from a limited amount of data.

New slide

1.5.4 Data Compression

Three basis principles of data compression for communication systems include:

Mathematically Lossless Compression This principle looks for mathematical coding schemes that reduce the *bits* required to represent a signal. For example, long runs of 0’s might be replaced by a shorter representation. This method of compression is used in computer file compression systems.

Lossy compression by removing redundant information This approach is often performed in a transform domain, such as the frequency domain. There might be many Fourier

(a) The digital versitile disc-audio (DVD-A) logo.

(b) The super-audio CD (SACD) logo.

Figure 1.22: High-quality audio formats.

coefficients that are small, and do not significantly contribute to the representation of the signal. If these small coefficients are not transmitted, then compression is achieved.

Lossless compression by linear prediction If it is possible to *predict* the current data sample from previous data samples, then it would not be necessary to transmit the current data symbol. Typically, however, the prediction is not completely accurate. However, by only transmitting the *difference* between the prediction and the actual value, which is typically a lot smaller than the actual value, then it turns out a fewer number of bits need to be transmitted, and thus compression achieved. The trick is to design a good *predictor*, and this is where statistical signal processing comes in handy.

1.5.5 Enhancement of Signals in Noise

High quality digital audio has in recent years dramatically raised expectations about sound quality. For example, high quality media such as:

- compact disc
- digital audio tape
- digital versitile disc-audio and super-audio CD.

Audio degradation is any undesirable modification to an audio signal occurring as the result of, or subsequent to, the recording process. Disturbances or distortions such as

1. background noise,
2. echoes and reverberation,
3. and media noise.

must be reduced to adequately low levels. Ideal restoration reconstructs the original sound exactly as would be received by transducers (microphone etc.,) in the absence of noise and acoustic distortion. Interest in historical material led to restoration of degraded sources including

1. wax cylinders recordings,

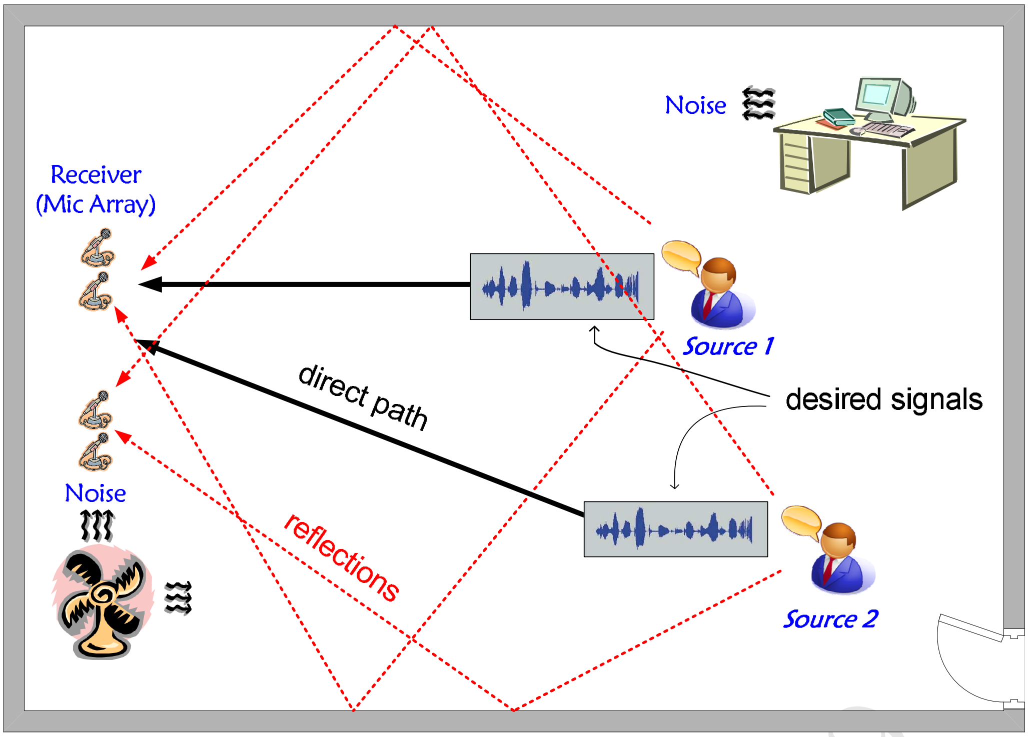
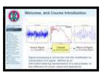


Figure 1.23: Passive source localisation and BSS.

2. disc recordings (78rpm, etc.),

3. and magnetic tape recordings.

Restoration is also required in contemporary digital recordings if distortion too intrusive. **Note** that noise present in recording environment, such as audience noise at a musical performance, considered part of *performance*. Statistical signal processing is required in such applications.



New slide

1.6 Passive and Active Target Localisation

This section presents a standard application in signal processing, namely passive target localisation. Active target localisation will be considered during the day as well, but this section will focus on the passive scenario. The aim of this section is to present, briefly, solutions to this problem, without restricting the notation used. If the mathematics is somewhat alien, then great, as the rest of this tutorial will explain the terms and concepts used here. An expanded version of this section, with a focus on acoustic source localisation, is included at the end of this handout.

A number of signal processing problems rely on knowledge of the desired source position, for example:

1. Tracking methods and target intent inference.

2. Mobile sensor node geometry.

3. Look-direction in beamforming techniques (for example in speech enhancement).

4. Camera steering for audio-visual BSS (including Robot Audition).

5. Speech diarisation.

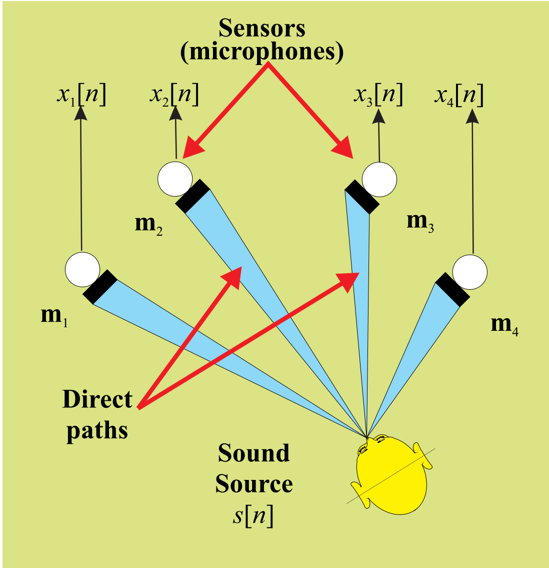
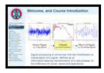


Figure 1.24: Ideal free-field model.

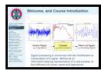
1.7 Passive Target Localisation Methodology



New slide

- In general, most passive target localisation (PTL) techniques rely on the fact that an impinging wavefront reaches one acoustic sensor before it reaches another.
- Most PTL algorithms are designed assuming there is no multipath or reverberation present, the *free-field assumption*.

1.7.1 Source Localization Strategies



New slide

Existing source localisation methods can loosely be divided into three generic strategies:

1. those based on maximising the steered response power (SRP) of a beamformer:
   - location estimate derived directly from a filtered, weighted, and sum version of the signal data received at the sensors;
2. techniques adopting high-resolution spectral estimation concepts:
   - any localisation scheme relying upon an application of the signal correlation matrix;
3. approaches employing time-difference of arrival (TDOA) information:
   - source locations calculated from a set of TDOA estimates measured across various combinations of sensors.

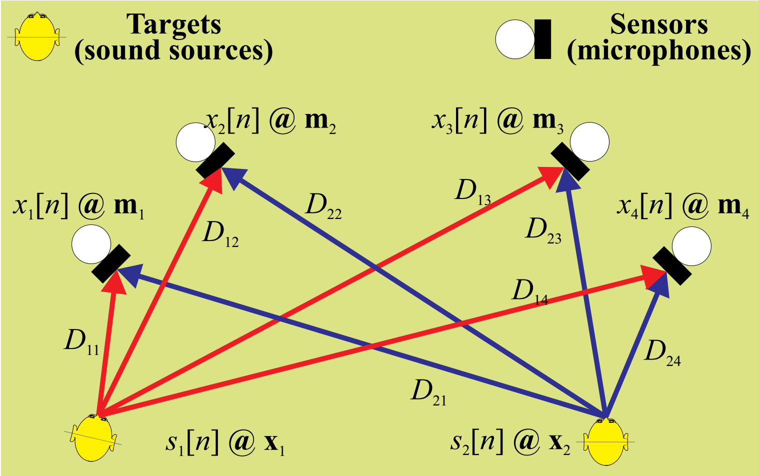
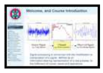


Figure 1.25: Geometry assuming a free-field model.

1.7.2 Geometric Layout



New slide

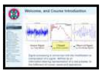
Suppose there is a:

- sensor array consisting of N nodes located at positions $\mathbf{m}_i \in \mathbb{R}^3$, for $i \in \{0, \dots, N-1\}$, and
- M talkers (or targets) at positions $\mathbf{x}_k \in \mathbb{R}^3$, for $k \in \{0, \dots, M-1\}$.

The TDOA between the sensor node at position $\mathbf{m}_i$ and $\mathbf{m}_j$ due to a source at $\mathbf{x}_k$ can be expressed as:

$$T(\mathbf{m}_i, \mathbf{m}_j, \mathbf{x}_k) \triangleq T_{ij}(\mathbf{x}_k) = \frac{|\mathbf{x}_k - \mathbf{m}_i| - |\mathbf{x}_k - \mathbf{m}_j|}{c} \tag{1.6}$$

where c is the speed of the impinging wavefront.



New slide

1.7.3 Ideal Free-field Model

- In an anechoic free-field environment, as depicted in Figure 1.24, the signal from source k, denoted $s_k(t)$, propagates to the i-th sensor at time t as:

  $$x_{ik}(t) = \alpha_{ik}\, s_k(t - \tau_{ik}) + b_{ik}(t) \tag{1.7}$$

  where $b_{ik}(t)$ denotes additive noise.

- Note that, in the frequency domain, this expression becomes:

  $$X_{ik}(\omega) = \alpha_{ik}\, S_k(\omega)\, e^{-j\omega\, \tau_{ik}} + B_{ik}(\omega) \tag{1.8}$$

  On the assumption of **geometrical wave propagation**, which assumes high frequencies, a point source of single frequency ω, at position $\mathbf{x}_k$ in free space, emits a pressure wave $P_{(\mathbf{x}_k, \mathbf{m}_i), t}(\omega)$ at time t and at position $\mathbf{m}_i$:

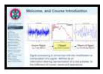
  $$P_{(\mathbf{x}_k, \mathbf{m}_i)}(\omega, t) = P_0 \frac{\exp\left[j\omega(r/c - t)\right]}{r} \tag{1.9}$$

  where $t \in \mathbb{R}$ is time, and $r = |\mathbf{x}_k - \mathbf{m}_i|$.

- The additive noise source is assumed to be uncorrelated with the source and noise sources at other sensors.
- The TDOA between the i-th and j-th sensor is given by:

$$\tau_{ijk} = \tau_{ik} - \tau_{jk} = T\left(\mathbf{m}_i,\, \mathbf{m}_j,\, \mathbf{x}_k\right) \tag{1.10}$$

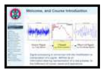
1.8 Indirect TDOA-based Methods


New slide

This is typically a two-step procedure in which:

- Typically, TDOAs are extracted using the generalised cross correlation (GCC) function, or an adaptive eigenvalue decomposition (AED) algorithm.
- A hypothesised spatial position of the target can be used to predict the expected TDOAs (or corresponding range) at the sensor.
- The error between the measured and hypothesised TDOAs is then minimised.
- Accurate and robust TDOA estimation is the key to the effectiveness of this class of PTL methods.
- An alternative way of viewing these solutions is to consider what **spatial positions** of the target could lead to the estimated TDOA.

1.8.1 Hyperbolic Least Squares Error Function


New slide

KEYPOINT! (Underlying Concept). Suppose that for each pair of sensors, i and j, a TDOA corresponding to source k is somehow estimated, and this is denoted by τ_{ijk}. One approach to ASL is to minimise the total error between the measured TDOAs and the TDOAs predicted by the geometry *given* an assumed target position.

- If a TDOA is estimated between two sensor nodes i and j, then the error between this and modelled TDOA is given by:

$$\epsilon_{ij}(\mathbf{x}_k) = \tau_{ijk} - T\left(\mathbf{m}_i,\, \mathbf{m}_j,\, \mathbf{x}_k\right) \tag{1.11}$$

  where the error is considered as a function of the source position $\mathbf{x}_k$.
- The total error as a function of target position

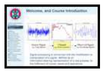
$$J(\mathbf{x}_k) = \sum_{i=1}^{N} \sum_{j \neq i=1}^{N} \epsilon_{ij}(\mathbf{x}_k) = \sum_{i=1}^{N} \sum_{j \neq i=1}^{N} \left(\tau_{ijk} - T\left(\mathbf{m}_i,\, \mathbf{m}_j,\, \mathbf{x}_k\right)\right)^2 \tag{1.12}$$

  where

$$T\left(\mathbf{m}_i,\, \mathbf{m}_j,\, \mathbf{x}_k\right) \triangleq T_{ij}\left(\mathbf{x}_k\right) = \frac{|\mathbf{x}_k - \mathbf{m}_i| - |\mathbf{x}_k - \mathbf{m}_j|}{c} \tag{1.13}$$

- Unfortunately, since $T\left(\mathbf{m}_i,\, \mathbf{m}_j,\, \mathbf{x}_k\right)$ is a nonlinear function of $\mathbf{x}_k$, the minimum least-squares estimate (LSE) does not possess a closed-form solution.

1.8.2 TDOA estimation methods



New slide

Two key methods for TDOA estimation are using the GCC function and the adaptive eigenvalue decomposition (AED) algorithm.

GCC algorithm most popular approach assuming an ideal free-field movel. It has the advantages that

- computationally efficient, and hence short decision delays;
- perform fairly well in moderately noisy and reverberant environments.

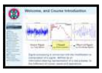
However, GCC-based methods

- fail when multipath is high;
- focus of current research is on combating the effect of multipath.

AED Algorithm Approaches the TDOA estimation approach from a different point of view from the *traditional* GCC method.

- adopts a multipath rather than free-field model;
- computationally more expensive than GCC;
- can fail when there are common-zeros in the channel.

Note that both methods assume that the signals received at the sensors arise as the result of a single source, and that if there are multiple sources, the signals will first need to be separated into different contributions of the individual sources.



New slide

1.8.2.1 GCC TDOA estimation

The GCC algorithm proposed by *Knapp and Carter* is the most widely used approach to TDOA estimation.

- The TDOA estimate between two microphones i and j is obtained as the time lag that maximises the cross-correlation between the filtered versions of the microphone outputs:

$$\hat{\tau_{ij}} = \arg\max_{\ell} r_{x_i x_j}[\ell] \tag{1.14}$$

  where the signal received at microphone i is given by $x_i[n]$, and where x_i should not be confused with the location of the source k, which is denoted by $\mathbf{x}_k = [x_k,\, y_k,\, z_k]^T$.

- The cross-correlation function is given by

$$r_{x_i x_j}[\ell] = \mathcal{F}^{-1}\left(\Psi_{x_1 x_2}\left(e^{j\omega T_s}\right)\right) \tag{1.15}$$

$$= \mathcal{F}^{-1}\left(\Phi\left(e^{j\omega T_s}\right) P_{x_1 x_2}\left(e^{j\omega T_s}\right)\right) \tag{1.16}$$

  where the cross-power spectral density (CPSD) is given by

$$P_{x_1 x_2}\left(e^{j\omega T_s}\right) = \mathbb{E}\left[X_1\left(e^{j\omega T_s}\right) X_2\left(e^{j\omega T_s}\right)\right] \tag{1.17}$$

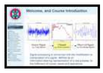
  The cross-power spectral density (CPSD) can be estimated in a variety of means. The choice of the filtering term or frequency domain weighting function, $\Phi\left(e^{j\omega T_s}\right)$, leads to a variety of different GCC methods for TDOA estimation.

- For the free-field model, it can be shown that:

$$\angle P_{x_i x_j}(\omega) = -j\omega\, T(\mathbf{m}_i,\, \mathbf{m}_j,\, \mathbf{x}_k) \tag{1.18}$$

In otherwords, all the TDOA information is conveyed in the phrase rather than the amplitude of the CPSD. This therefore suggests that the weighting function can be chosen to remove the amplitude information.

1.8.2.2 GCC Processors



New slide

The most common choices for the GCC weighting term are listed in the table below. In particular, the phase transform (PHAT) is considered in detail.

Processor Name	Frequency Function
Cross Correlation	1
PHAT	$\dfrac{1}{\lvert P_{x_1x_2}(e^{j\omega T_s})\rvert}$
Roth Impulse Response	$\dfrac{1}{P_{x_1x_1}(e^{j\omega T_s})}$ or $\dfrac{1}{P_{x_2x_2}(e^{j\omega T_s})}$
SCOT	$\dfrac{1}{\sqrt{P_{x_1x_1}(e^{j\omega T_s})\, P_{x_2x_2}(e^{j\omega T_s})}}$
Eckart	$\dfrac{P_{s_1s_1}(e^{j\omega T_s})}{P_{n_1n_1}(e^{j\omega T_s})\, P_{n_2n_2}(e^{j\omega T_s})}$
Hannon-Thomson or ML	$\dfrac{\lvert\gamma_{x_1x_2}(e^{j\omega T_s})\rvert^2}{\lvert P_{x_1x_2}(e^{j\omega T_s})\rvert\left(1-\lvert\gamma_{x_1x_2}(e^{j\omega T_s})\rvert^2\right)}$

where $\gamma_{x_1x_2}(e^{j\omega T_s})$ is the normalised CPSD or **coherence function** is given by

$$\gamma_{x_1x_2}(e^{j\omega T_s}) = \frac{P_{x_1x_2}(e^{j\omega T_s})}{\sqrt{P_{x_1x_1}(e^{j\omega T_s})\, P_{x_2x_2}(e^{j\omega T_s})}} \tag{1.19}$$

The PHAT-GCC approach can be written as:

$$r_{x_i x_j}[\ell] = \int_{-\frac{\pi}{T_s}}^{\frac{\pi}{T_s}} \Phi(e^{j\omega T_s})\, P_{x_1x_2}(e^{j\omega T_s})\, e^{j\ell\omega T}\, d\omega \tag{1.20}$$

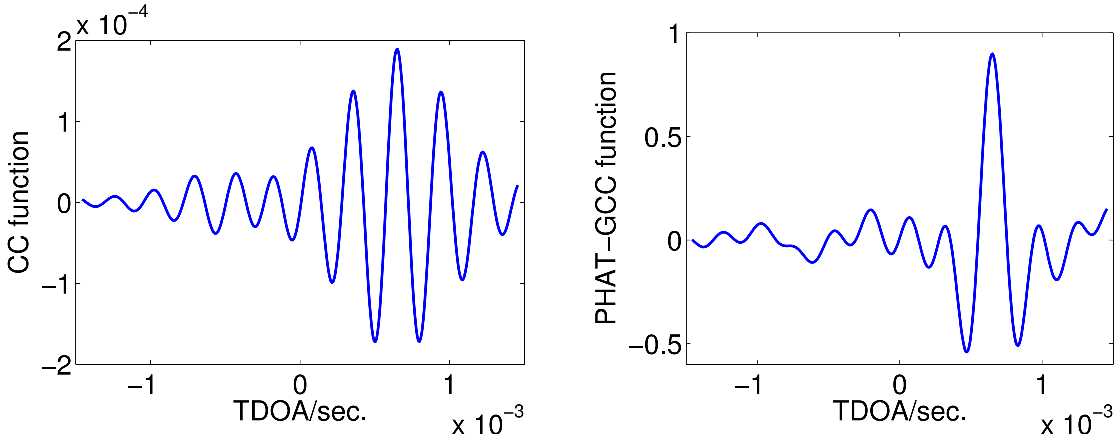
$$= \int_{-\frac{\pi}{T_s}}^{\frac{\pi}{T_s}} \frac{1}{\lvert P_{x_1x_2}(e^{j\omega T_s})\rvert} \lvert P_{x_1x_2}(e^{j\omega T_s})\rvert e^{j\angle P_{x_1x_2}(e^{j\omega T_s})}\, e^{j\ell\omega T}\, d\omega \tag{1.21}$$

$$= \int_{-\frac{\pi}{T_s}}^{\frac{\pi}{T_s}} e^{j\left(\ell\omega T + \angle P_{x_1x_2}(e^{j\omega T_s})\right)}\, d\omega \tag{1.22}$$

$$= \delta\big(\ell\, T_s + \angle P_{x_1x_2}(e^{j\omega T_s})\big) \tag{1.23}$$

$$= \delta(\ell\, T_s - T(\mathbf{m}_i,\, \mathbf{m}_j,\, \mathbf{x}_k)) \tag{1.24}$$

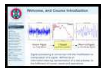
- In the absence of reverberation, the GCC-PHAT (GCC-PHAT) algorithm gives an impulse at a lag given by the TDOA divided by the sampling period.



(a) Cross-correlation function. (b) GCC-PHAT function

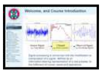
Figure 1.26: Normal cross-correlation and GCC-PHAT functions for a frame of speech.

1.9 Direct Localisation Methods



New slide

- Direct localisation methods have the advantage that the relationship between the measurement and the state is linear.
- However, extracting the position measurement requires a multi-dimensional search over the state space and is usually computationally expensive.



New slide

1.9.1 Steered Response Power Function

KEYPOINT! (Underlying Concept). The steered beamformer (SBF) or SRP function is a measure of correlation across *all pairs* of microphone signals for a set of relative delays that arise from a hypothesised source location.

The frequency domain **delay-and-sum beamformer** steered to a spatial position $\hat{\mathbf{x}}_k$ such that $\hat{\tau}_{pk} = |\hat{\mathbf{x}} - \mathbf{m}_p|$, using the notation in Equation 7.8, is given by:

$$S(\hat{\mathbf{x}}) = \int_\Omega \left| \sum_{p=1}^{N} W_p\left(e^{j\omega T_s}\right) X_p\left(e^{j\omega T_s}\right) e^{j\omega\,\hat{\tau}_{pk}} \right|^2 d\omega \tag{1.25}$$

Expanding, rearranging the order of integration and summation, taking expectations of both sides and setting $\Phi_{pq}\left(e^{j\omega T_s}\right) = W_p\left(e^{j\omega T_s}\right) W_q^*\left(e^{j\omega T_s}\right)$ gives

$$\mathbb{E}\left[S(\hat{\mathbf{x}})\right] = \sum_{p=1}^{N}\sum_{q=1}^{N} r_{x_i x_j}[\hat{\tau}_{pqk}] \tag{1.26}$$

$$\equiv \sum_{p=1}^{N}\sum_{q=1}^{N} r_{x_i x_j}\left[\frac{|\mathbf{x}_k - \mathbf{m}_i| - |\mathbf{x}_k - \mathbf{m}_j|}{c}\right] \tag{1.27}$$

In other words, the SRP is the sum of all possible pairwise GCC functions evaluated at the time delays hypothesised by the target position.

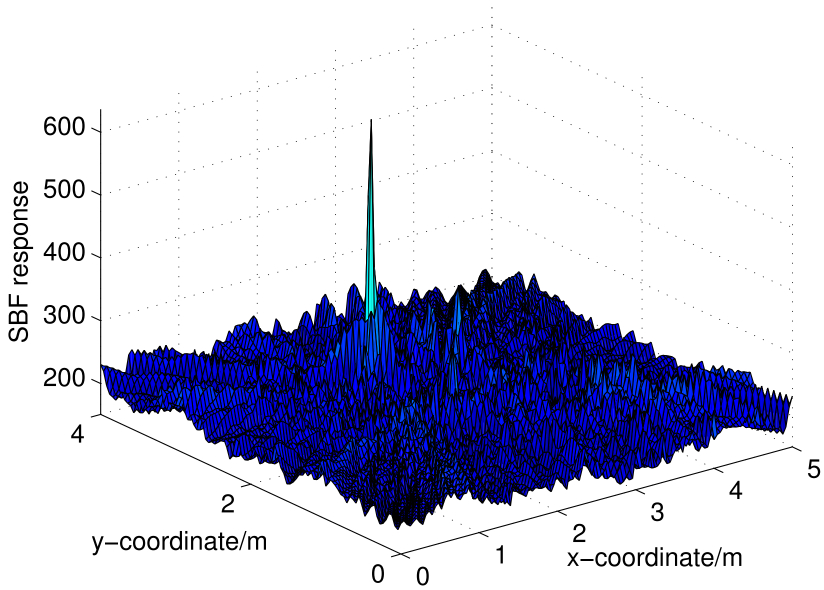


Figure 1.27: SBF response from a frame of speech signal. The integration frequency range is 300 to 3500 Hz (see Equation 7.84). The true source position is at $[2.0, 2.5]m$. The grid density is set to 40 mm.

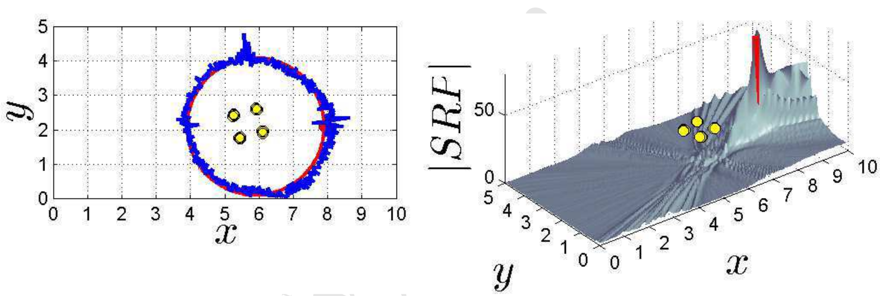
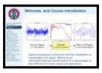


Figure 1.28: An example video showing the SBF changing as the source location moves.



New slide

1.9.2 Conclusions

To fully appreciate the algorithms in PTL, we need:

1. Signal analysis in time and frequency domain.
2. Least Squares Estimation Theory.
3. Expectations and frequency-domain statistical analysis.
4. Correlation and power-spectral density theory.
5. And, of course, all the theory to explain the above!

2

Review of Basic Probability Theory

All knowledge degenerates into probability.

David Hume

> This handout gives a review of the fundamentals of probability theory. The idea is to motivate the definitions of cumulative distribution functions (cdfs) and probability density functions (pdfs) in the next handouts.

2.1 Introduction

New slide

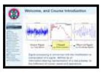
The theory of probability deals with averages of mass phenomena occurring sequentially or simultaneously; in signal processing and communications, this might include radar detection, signal detection, anomaly detection, parameter estimation, and so forth.

How does one start considering the notion and meaning of probability, and how can it be extended to model signals? To address this, it is first important to consider the probability of individual events. It has been *observed* in many fields that certain averages approach a constant value as the number of observations increases, and this value remains the same if the averages are evaluated over any subsequence (of observations) specified before the experiment is performed. In a coin experiment, for example, the percentage of heads approaches 0.5 or some other constant, and the same average is obtained if every fourth, sixth, or arbitrary selection of tosses is chosen. Note that the notion of an average is not in-itself a probabilistic term.

The purpose of the theory of probability is to describe and predict these averages in terms of probabilities of events. The probability of an event A is a number $\Pr(A)$ assigned to this event. This number *could* be interpreted as follows:

> If an experiment is performed n times, and the event A occurs n_A times, then with a *high degree of certainty*, the relative frequency n_A/n is *close to* $\Pr(A)$, such that:
>
> $$\Pr(A) \approx \frac{n_A}{n} \tag{2.1}$$
>
> provided that n is *sufficiently large*.

Note that this interpretation and the language used is all very imprecise, and phrases such as *high degree of certainty*, *close to*, and *sufficiently large* has no clear meaning. These terms will be more precisely defined as concepts are introduced throughout this course.



New slide

2.2 Classical Definition of Probability

For several centuries, the theory of probability was based on the *classical definition*, which states that the probability $\Pr(A)$ of an event A is determine *a priori* without actual experimentation. It is given by the ratio:

$$\Pr(A) = \frac{N_A}{N} \tag{2.2}$$

where:

- N is the total number of outcomes,
- and N_A is the total number of outcomes that are favourable to the event A, provided that *all outcomes are equally probable*.

This definition, however, has some difficulties when the number of possible outcomes is infinite, as illustrated in the following example in Section 2.2.1.

New slide

2.2.1 Bertrand's Paradox

Consider a circle C of radius r; what is the probability p that the length ℓ of a *randomly selected* cord AB is greater than the length, $r\sqrt{3}$, of the inscribed equilateral triangle?

> **KEYPOINT! (Recalling Geometry!).** To fully appreciate this problem, it is perhaps worth being aware of the geometry of this problem. The idea of the geometry is to keep simple geometric shapes, rather than to play on some obscure geometric properties. Therefore, note that if three tangents to a circle of radius $r/2$ are drawn at angular intervals of $120\,\text{degs}$, then the resulting equilateral triangle fits inside a larger circle of radius r, as shown in Figure 2.1. The length of the sides of one of this equilateral triangle is $r\sqrt{3}$.

Using the classical definition of probability, three reasonable solutions can be obtained:

Sidebar 1 The Venice Water-Taxi Problem

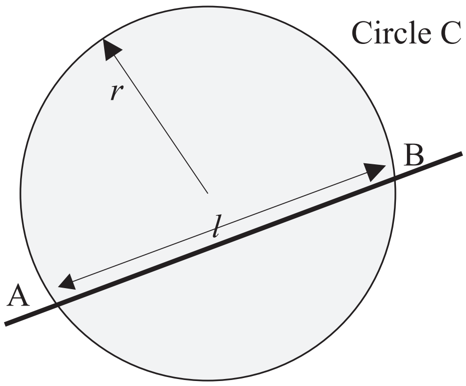
Understanding probability and statistics helps understand simple, but important, questions related to estimating the parameters of a sampling distribution from a small sample size.

On a recent trip to Venice, it was observed that the water taxis were numbered in sequential order from number 1 up-wards (a water-taxi with the number 1 on the side was observed, and only positive integer valued taxi designations).

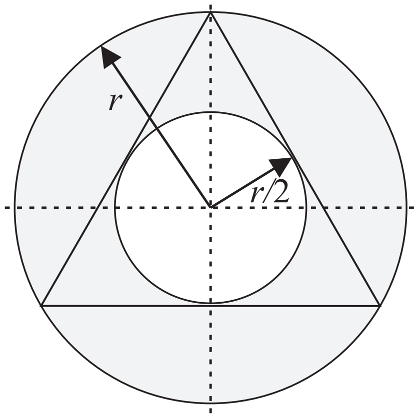

Assuming that all taxis are in service, suppose we wanted to guess the number N of water taxis in Venice, based purely on the taxi numbers observed. Let's assume we observed a taxi with the number 304 on the side. What is our best guess of N?

The solution will be discussed in detail in Chapter 5, but now is a good time to think about it in advance of learning the techniques that will help us answer the question. Moreover, suppose we observe more taxis, perhaps with the numbers 157, 202, 11, 248; how will our estimate change?


This problem might seem rather academic, but has actually in the past been far from it, as discussed in Chapter 5 as well.

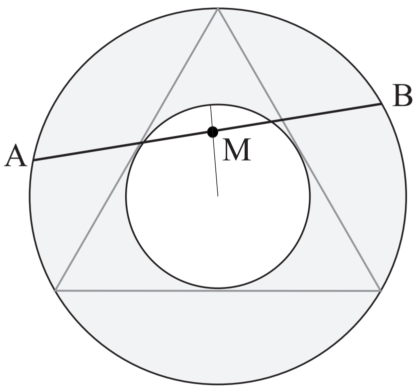


(a) Basic problem: line placed across circle at random, resulting in the *cord* AB.

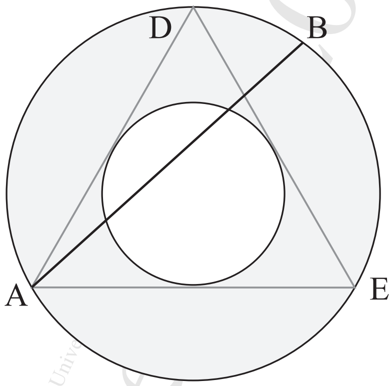


(b) The problem definition is setup so that the nice geometrical properties of the equilateral triangle can be used.

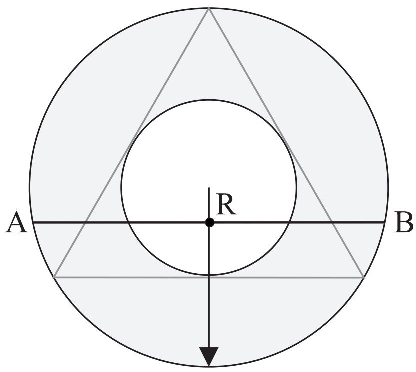
Figure 2.1: Bertrand's paradox, problem definition.



(a) The *midpoint method.*



(b) The *endpoint method.*



(c) The *radius method.*

Figure 2.2: Different selection methods.

1. In the **random midpoints** method, a cord is selected by choosing a point M anywhere in the full circle, and two end-points A and B on the circumference of the circle, such that the resulting chord AB through these chosen points has M as its midpoint. This is shown graphically in Figure 2.2a.

   It is reasonable, therefore, to consider as *favourable outcomes* all points inside the inner-circle of radius $r/2$, and to consider *all possible outcomes* as points inside the outer-circle of radius r. Therefore, using as a measure of these outcomes the corresponding areas, it follows that:

$$p = \frac{\pi \left(\frac{r}{2}\right)^2}{\pi r^2} = \frac{1}{4} \tag{2.3}$$

2. In the **random endpoints** method, consider selecting two random points on the circumference of the (outer) circle, A and B, and drawing a chord between them. This is shown in Figure 2.2b, where the point A has been drawn to coincide with the particular triangle drawn. If B lies on the arc between the two other vertices, D and E, of the triangle whose first vertex coincides with A, then AB will be longer than the length of the side of the triangle.

   The *favourable outcomes* are now the points on this arc, and since the angle of the arc DE is $\frac{2\pi}{3}$ radians, a measure of this outcome is the arc length $\frac{2\pi r}{3}$. Moreover, the total outcomes are all the points on the circumference of the main circle, and therefore it follows:

$$p = \frac{\frac{2\pi r}{3}}{2\pi r} = \frac{1}{3} \tag{2.4}$$

3. Finally, in the **random radius method**, a radius of the circle is chosen at random, and a point on the radius is chosen at random. The chord AB is constructed as a line perpendicular to the chosen radius through the chosen point. The construction of this chord is shown in Figure 2.2c.

   The *favourable outcomes* are the points on the radius that lie *inside* of the inner-circle, or a measure of this outcome is given by the diameter of the inner-circle, r. The total outcomes are the points on the diameter of the outer-circle, and a measure of that respective length is $2r$. Therefore, the probability is given by

$$p = \frac{r}{2r} = \frac{1}{2} \tag{2.5}$$

There are thus three different but reasonable solutions to the same problem. Which one is valid?

2.2.2 Using the Classical Definition

The difficulty with the classical definition in Equation 2.2, as seen in Bertrand's Paradox, is in determining N and N_A.

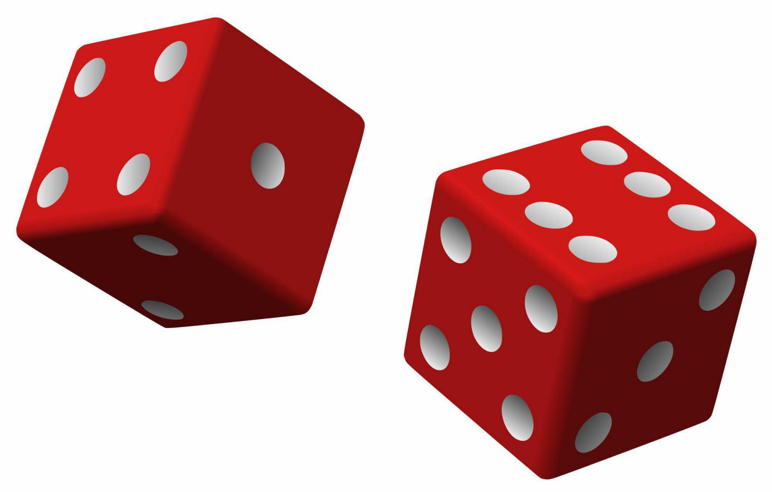


Figure 2.3: Two red dice: https://commons.wikimedia.org/wiki/File:Two_red_dice_01.sv

Example 2.1 (Rolling two dice). Two dice are rolled (see Figure 2.3); find the probability, p, that the sum of the numbers shown equals 7. Consider three possibilities:

1. The *possible outcomes* total 11 which are the sums $\{2, 3, \dots, 12\}$. Of these, only one (the sum 7) is favourable. Hence, $p = \frac{1}{11}$.

   This is, of course, wrong, and the reason is that each of the 11 possible outcomes are *not* equally probable.

2. Similarly, writing down the possible pairs of shown numbers, without distinguishing between the first and second die. There are then 21 pairs, $(1, 1)$, $(1, 2)$, $\dots$, $(1, 6)$, $(2, 1)$, $\dots$, $(6, 6)$, of which there are three favourable pairs $(3, 4)$, $(5, 2)$ and $(6, 1)$. However, again, the pairs $(3, 4)$ and $(6, 6)$, for example, are not equally likely.

3. Therefore, to count all possible outcomes which are equally probable, it is necessary to could all pairs of numbers distinguishing between the first and second die. This will give the correct probability.

New slide

2.2.3 Difficulties with the Classical Definition

The classical definition in Equation 2.2 can be questioned on several grounds, namely:

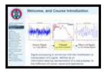
1. The term **equally probable** in the definition of probability is making use of a concept still to be defined!

2. The definition can only be applied to a limited class of problems.

   In the die experiment, for example, it is applicable only if the six faces have the same probability. If the die is loaded and the probability of a "4" equals 0.2, say, then this cannot be determined from the classical ratio in Equation 2.2.

3. If the number of possible outcomes is infinite, then some other measure of infinity for determining the classical probability ratio in Equation 2.2 is needed, such as length, or area. This leads to difficulties, as discussed in Bertrand's paradox.

2.3 Axiomatic Definition



New slide

The axiomatic approach to probability is based on the following three postulates and *on nothing else*:

1. The probability $\Pr(A)$ of an event A is a non-negative number assigned to this event:

$$\Pr(A) \geq 0 \tag{2.6}$$

2. Defining the **certain event**, S, as the event that occurs in every trial, then the probability of the certain event equals 1, such that:

$$\Pr(S) = 1 \tag{2.7}$$

3. If the events A and B are **mutually exclusive**, then the probability of one event or the other occurring separately is:

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) \tag{2.8}$$

   or more generally, if $A_1, A_2, \ldots$ is a collection of disjoint events, such that $A_i \cap A_j = \emptyset$ for all pairs i, j satisfying $i \neq j$, then:

$$\Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \Pr(A_i) \tag{2.9}$$

   Note that Equation 2.9 does not directly follow from Equation 2.8, even though it may appear to. Dealing with infinitely many sets requires further insight, and here the result of Equation 2.9 is actually an additional condition known as the **axiom of infinite additivity**.

These axioms can be formalised by defining measures and fields as appropriate, but the level of detail is beyond this course.

These axioms, once formalised, are known as the **Kolmogorov Axioms**, named after the Russian mathematician. Note that an alternative approach to deriving the laws of probability theory from a certain set of postulates was developed by Cox. However, this won't be considered in this course.

2.3.1 Set Theory

New slide

Since the classical definition of probability details in total number of outcomes, as well as events, it is necessary to utilise the mathematical language of sets to formulise precise definitions.

A **set** is a collection of objects called **elements**. For example, "*car*, *apple*, *pencil*" is a set with three elements whose elements are a car, an apple, and a pencil. The set "*heads*, *tails*" has two elements, while the set "1, 2, 3, 5", has four. It is assumed that most readers will have come across **set theory** to some extent, and therefore, it will be used throughout the document as and when needed.

Some basic notation, however, includes the following:

Unions and Intersections Unions and intersections are commutative, associative, and distributive, such that:

$$A \cup B = B \cup A, \quad (A \cup B) \cup C = A \cup (B \cup C) \tag{2.10}$$

$$AB = BA, \quad (AB)C = A(BC), \quad A(B \cup C) = AB \cup AC \tag{2.11}$$

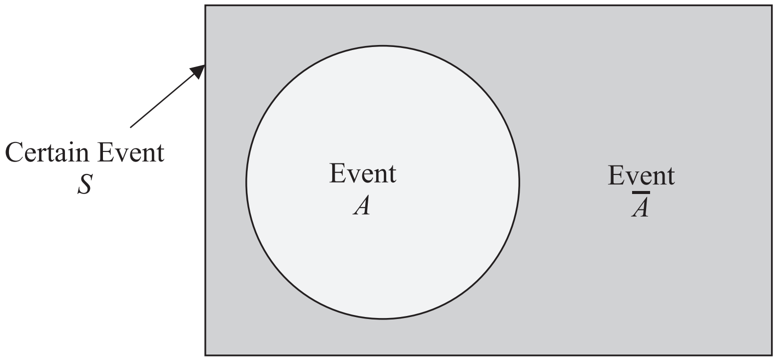


Figure 2.4: The complement $\overline{A}$ of $A \subset S$ is the set of all elements of S not in A.

Complements The complement $\overline{A}$ of a set $A \subset S$ is the set consisting of all elements of S that are not in A. Note that:

$$A \cup \overline{A} = S \quad \text{and} \quad A \cap \overline{A} \equiv A\overline{A} = \{\emptyset\} \tag{2.12}$$

This is shown graphically using a Venn diagram, as shown in Figure 2.4.

Partitions A partition U of a set S is a collection of mutually exclusive subsets A_i of S whose union equations S, such that:

$$\bigcup_{i=1}^{\infty} A_i = S, \quad A_i \cap A_j = \{\emptyset\}, \quad i \neq j \quad \Rightarrow \quad U = [A_1, \dots, A_n] \tag{2.13}$$

De Morgan's Law Using Venn diagrams, it is relatively straightforward to show

$$\overline{A \cup B} = \overline{A} \cap \overline{B} \equiv \overline{A}\,\overline{B} \quad \text{and} \quad \overline{A \cap B} \equiv \overline{AB} = \overline{A} \cup \overline{B} \tag{2.14}$$

As an application of this, note that:

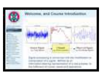
$$\overline{A \cup BC} = \overline{A}\,\overline{BC} = \overline{A}\left(\overline{B} \cup \overline{C}\right) \tag{2.15}$$

$$= \left(\overline{A}\,\overline{B}\right) \cup \left(\overline{A}\,\overline{C}\right) \tag{2.16}$$

$$= \overline{A \cup B} \cup \overline{A \cup C} \tag{2.17}$$

$$\Rightarrow \quad A \cup BC = (A \cup B)(A \cup C) \tag{2.18}$$

This result can easily be derived by using Venn diagrams, and it is worth checking this result yourself. This latter identity will also be used later in Section 2.3.2.



New slide

2.3.2 Properties of Axiomatic Probability

Some simple consequences of the definition of probability defined in Section 2.3 follow immediately:

Impossible Event The probability of the impossible event is 0, and therefore:

$$\Pr(\emptyset) = 0 \tag{2.19}$$

Complements Since $A \cup \overline{A} = S$ and $A\overline{A} = \{\emptyset\}$, then using Equation 2.8, $\Pr\left(A \cup \overline{A}\right) = \Pr(A) + \Pr\left(\overline{A}\right) = \Pr(S) = 1$, such that:

$$\Pr\left(\overline{A}\right) = 1 - \Pr(A) \tag{2.20}$$

Sum Rule The **addition law of probability** or the **sum rule** for any two events A and B is given by:

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B) \tag{2.21}$$

Example 2.2 (Proof of the Sum Rule). Prove the result in Equation 2.21.

SOLUTION. To prove this, separately write $A \cup B$ and B as the union of two mutually exclusive events (using Equation 2.18 and the fact $A \cup \overline{A} = S$ and $S\,B = B$).

- First, note that

$$A \cup B = \left(A \cup \overline{A}\right)(A \cup B) = A \cup \left(\overline{A}\,B\right) \tag{2.22}$$

  and that since $A\,\left(\overline{A}\,B\right) = \left(A\,\overline{A}\right)B = \{\emptyset\}B = \{\emptyset\}$, then A and $\overline{A}\,B$ are mutually exclusive events.

- Second, note that:

$$B = \left(A \cup \overline{A}\right)B = (A\,B) \cup \left(\overline{A}\,B\right) \tag{2.23}$$

  and that $(A\,B) \cap \left(\overline{A}\,B\right) = A\,\overline{A}\,B = \{\emptyset\}\,B = \{\emptyset\}$ and are therefore mutually exclusive events.

Using these two disjoint unions, then:

$$\Pr(A \cup B) = \Pr\left(A \cup \left(\overline{A}\,B\right)\right) = \Pr(A) + \Pr\left(\overline{A}\,B\right) \tag{2.24}$$

$$\Pr(B) = \Pr\left((A\,B) \cup \left(\overline{A}\,B\right)\right) = \Pr(A\,B) + \Pr\left(\overline{A}\,B\right) \tag{2.25}$$

Eliminating $\Pr\left(\overline{A}\,B\right)$ by subtracting these equations gives the desired result:

$$\Pr(A \cup B) - \Pr(B) = \Pr\left(A \cup \left(\overline{A}\,B\right)\right) = \Pr(A) - \Pr(A\,B) \tag{2.26}$$

□

Example 2.3 (Sum Rule). Let A and B be events with probabilities $\Pr(A) = 3/4$ and $\Pr(B) = 1/3$. Show that $1/12 \le \Pr(A\,B) \le 1/3$.

SOLUTION. Using the sum rule, that:

$$\Pr(A\,B) = \Pr(A) + \Pr(B) - \Pr(A \cup B) \ge \Pr(A) + \Pr(B) - 1 = \frac{1}{12} \tag{2.27}$$

□

which is the case when the whole **sample space** is covered by the two events. The second bound occurs since $A \cap B \subset B$ and similarly $A \cap B \subset A$, where $\subset$ denotes subset. Therefore, it can be deduced $\Pr(A\,B) \le \min\{\Pr(A), \Pr(B)\} = 1/3$.

2.3.3 Countable Spaces

If the **certain event**, S, consists of N outcomes, and N is a finite number, then the probabilities of all events can be expressed in terms of the probabilities $\Pr(\zeta_i) = p_i$ of the elementary events $\{\zeta_i\}$.

Example 2.4 (Cups and Saucers). Six cups and saucers come in pairs: there are two cups and saucers which are RED, two which are GREEN, and two which are BLUE. If the cups are placed randomly onto the saucers (one each), find the probability that no cup is upon a saucer of the same pattern.

SOLUTION.

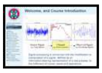
- Lay the saucers in order, say as $RRGGBB$.
- The cups may be arranged in $6!$ ways, but since each pair of a given colour may be switched without changing the appearance, there are $6!/(2!)^3 = 90$ distinct arrangements.

  By assumption, each of these are equally likely.
- The arrangements in which cups never match their saucers are:

$$\begin{array}{cccc} \underline{GG}BBRR, & \underline{GB}RBGR, & \underline{BG}RBGR, & \underline{BB}RRGG \\ & \underline{GB}BRGR, & \underline{BG}BRGR & \\ & \underline{GB}RBRG, & \underline{BG}RBRG & \\ & \underline{GB}BRGR, & \underline{BG}BRGR & \end{array} \tag{2.28}$$

□

- Hence, the required probability is $^{10}/_{90} = 1/9$.



New slide

2.3.4 The Real Line

If the **certain event**, S, consists of a non-countable infinity of elements, then its probabilities cannot be determined in terms of the probabilities of elementary events. This is the case if S is the set of points in an n-dimensional space.

Suppose that S is the set of all real numbers. Its subsets can be considered as sets of points on the real line. To construct a probability space on the real line, consider events as intervals $x_1 < x \leq x_2$, and their countable unions and intersections.

To complete the specification of probabilities for this set, it suffices to assign probabilities to the events $\{x \leq x_i\}$.

This notion leads to **cumulative distribution functions (cdfs)** and **probability density functions (pdfs)** in the next handout.

New slide

2.4 Conditional Probability

To introduce conditional probability, consider the discussion about proportions in Section 2.1. If an experiment is repeated n times, and on each occasion the occurrences or non-occurrences of two events A and B are observed. Suppose that only those outcomes for which B occurs are considered, and all other experiments are disregarded.

In this smaller collection of trials, the proportion of times that A occurs, given that B has occurred, is:

$$\Pr\left(A \mid B\right) \approx \frac{n_{AB}}{n_B} = \frac{^{n_{AB}}/_n}{^{n_B}/_n} = \frac{\Pr\left(AB\right)}{\Pr\left(B\right)} \tag{2.29}$$

provided that n is sufficiently large.

The **conditional probability** of an event A assuming another event B, denoted by $\Pr\left(A \,\middle|\, B\right)$, is defined by the ratio:

$$\Pr\left(A \,\middle|\, B\right) = \frac{\Pr\left(A \cap B\right)}{\Pr\left(B\right)} \tag{2.30}$$

It can be shown that this definition satisfies the **Kolmogorov Axioms**.

Example 2.5 (Two Children). A family has two children. What is the probability that both are boys, given that at least one is a boy?

SOLUTION. The younger and older children may each be male or female, and it is assumed that each is equally likely.

There are four possibilities for the gender of the children, namely:

$$S = \{GG,\ GB,\ BG,\ BB\} \tag{2.31}$$

where the four possibilities are equally probable:

$$\Pr\left(GG\right) = \Pr\left(GB\right) = \Pr\left(BG\right) = \Pr\left(BB\right) = \frac{1}{4} \tag{2.32}$$

The subset of S which contains the possibilities of one child being a boy is at $S_B = \{GB,\ BG,\ BB\}$, and therefore the conditional probability:

$$\Pr\left(BB \,\middle|\, S_B\right) = \frac{\Pr\left(BB \cap (GB \cup BG \cup BB)\right)}{\Pr\left(S_B\right)} \tag{2.33}$$

Note that $\{BB \cap (GB \cup BG \cup BB)\} = \{BB\}$, and that $\Pr\left(S_B\right) = 1 - \Pr\left(S_B\right) = 1 - \Pr\left(GG\right) = \frac{3}{4}$. Therefore:

$$\Pr\left(BB \,\middle|\, S_B\right) = \frac{\Pr\left(BB\right)}{1 - \Pr\left(GG\right)} = \frac{1/4}{3/4} = \frac{1}{3} \tag{2.34}$$

□

Note that the question is completely different if it were *what is the probability that both are boys, given that the youngest child is a boy*, in which case the solution is $1/2$. This is since information has been provided about one of the children, thereby distinguishing between the children.

Example 2.6 (Two Children (Variant)). A family has two children. One of the children is a boy born in an *even* month, where even months are defined as *Feburary, April, June, August, October, and December*, while odd months are defined as *January, March, May, July, September, and November*. What is the probability that both are boys?

SOLUTION. The younger and older children may each be male or female, and it is assumed that each is equally likely. Moreover, the month in which each child is born is assumed to be equally likely. Denoting the first child as C_1, and the second by C_2, there are 16 different but equally likely possibilities, which are denoted given by:

C_1		C_2		Outcome	
Gender	Month	Gender	Month	Relevant?	Desired?
B	O	B	O		
B	O	B	E	✓	✓
B	E	B	O	✓	✓
B	E	B	E	✓	✓
G	O	B	O		
G	O	B	E	✓	
G	E	B	O		
G	E	B	E	✓	
B	O	G	O		
B	O	G	E		
B	E	G	O	✓	
B	E	G	E	✓	
G	O	G	O		
G	O	G	E		
G	E	G	O		
G	E	G	E		
Count				7	3

□

Therefore, the number of favourable outcomes to the question in hand is $3/7 = 0.428$, which is getting closer to one half than a third.

The example in Example 2.5 might seem a little abstract to signal processing, but there are other ways of phrasing exactly the same problem. Using an example taken from [Therrien:2011], it could be phrased as follows:

> A compact disc (CD) selected from the *bins* at Simon's Surplus are as likely to be good as they are bad. Simon decides to sell these CDs in packages of two, but guarantees that in each package, at least one CD will be good. What is the probability that when you buy a single package, you get two good CDs?

It should be apparent that this is the same problem as in Example 2.5. One further problem to consider is given below in Example 2.7.

A further example discussed in the lectures covers mobile phones; a company sells mobile phones in boxes, and are equally likely to be broken (B) or working (W). You are given two boxes and told that in one of the boxes there is a working phone. What is the probability that the other box also contains a working phone? Suppose now that all phones are manufactured by four companies: A, E, N, and S. You are told that one of the boxes contains a working phone manufactured by company S. What is the probability that the other box contains a working phone?

Finally, to extend the discussion further, suppose all the phones are made between the years 1997 and 2016, and by the four companies above. One of the boxes contains a working phone made in 2007 by manufacturer A. What is the probability the other box contains a working phone? It should be apparent that by giving more information about one of the phones, the probability of the other box containing a working phone approaches a half.

Example 2.7 (Prisoner's Paradox). Three prisoners, A, B and C, are in separate cells and sentenced to death. The governor has selected one of them at random to be pardoned. The warden knows which one is pardoned, but is not allowed to tell. Prisoner A begs the warden to let him know the identity of one of the *others* who is going to be executed.

> *If B is to be pardoned, give me C's name, and vice-versa. And if I'm to be pardoned, flip a coin to decide whether to name B or C.*

The warden tells A that B is to be executed. Prisoner A is pleased because he believes that his probability of surviving has gone up from $1/3$ to $1/2$, as it is now between him and C. Prisoner A secretly tells C the news, who is also pleased, because he reasons that A still has a chance of $1/3$ to be the pardoned one, but his chance has gone up to $2/3$. What is the correct answer?

3

Scalar Random Variables

This handout introduces the concept of a random variable, its probabilistic description in terms of pdfs and cdfs, and characteristic features such as mean, variance, and other moments. It covers the probability transformation rule and characteristic functions.

3.1 Abstract

- Deterministic signals are interesting from an analytical perspective since their *signal value* or *amplitude* are uniquely and completely specified by a functional form, albeit that function might be very complicated. Thus, a deterministic signal is some function of time: $x = x(t)$.
- In practice, this precise description cannot be obtained for real-world signals and, moreover, it can be argued philosophically that real-world signals are not deterministic but, rather, they are inherently random or *stochastic* in nature.
- Although random signals evolve in time stochastically, their average properties are often deterministic, and thus can be specified by an explicit functional form.
- This part of the course looks at the properties of stochastic processes, both in terms of an exact probabilistic description, and also characteristic features such as mean, variance, and other moments.

New slide

3.2 Definition Random Variables

A **random variable (RV)** $X(\zeta)$ is a mapping that assigns a real number $X \in (-\infty, \infty)$ to every outcome ζ from an abstract probability space. This mapping from ζ to X should satisfy the following two conditions:

1. the interval $\{X(\zeta) \leq x\}$ is an event in the abstract probability space for every $x \in \mathbb{R}$;
2. $\Pr(X(\zeta) = \infty) = 0$ and $\Pr(X(\zeta) = -\infty) = 0$.

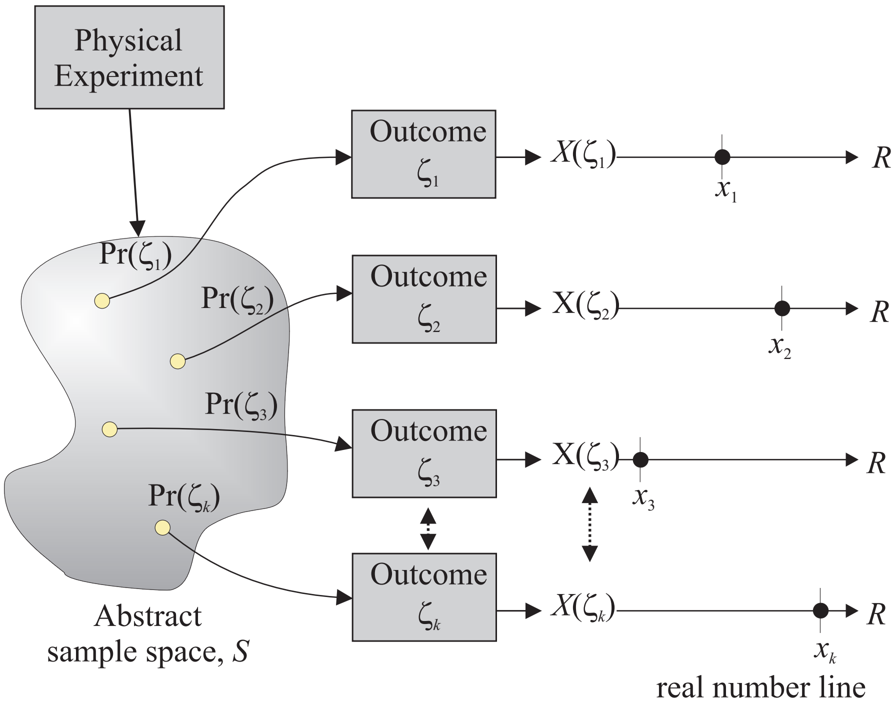


Figure 3.1: A graphical representation of a random variable.

The second condition states that, although X is allowed to take the values $x = \pm\infty$, the outcomes form a set with zero probability.

KEYPOINT! (Nature of Outcomes). Note that the outcomes of events are not necessarily numbers themselves, although they should be distinct in nature. Hence, examples of outcomes might be:

- outcomes of tossing coins (head/tails); card drawn from a deck (King, Queen, 8-of-Hearts);
- characters or words (A-Z); symbols used in deoxyribonucleic acid (DNA) sequencing (A, T, G, C);
- a numerical result, such as the number rolled on a die.

A more graphical representation of a discrete RV is shown in Figure 3.1. In this model, a physical experiment can lead to a number of possible events representing the outcomes of the experiment. These outcomes may be values, or they may be symbols, or some other representation of the event. Each outcome (or event), ζ_k, has a probability $\Pr(\zeta_k)$ assigned to it. Each outcome ζ_k then a real number assigned to that outcome, x_k. The RV is then defined as the collection of these three values; an outcome index, the probability of the outcome, and the real value assigned to that outcome, thus $X(\zeta) = \{\zeta_k, \Pr(\zeta_k), x_k$.

A more specific example is shown in Figure 3.2 in which the **experiment** is that of rolling a die, the **outcomes** are the colors of the dies, each **event** is simply each **outcome**, and the specific user-defined values assigned are the numbers shown.

Example 3.1 (Rolling die). Consider rolling a die, with six outcomes $\{\zeta_i, i \in \{1, \dots, 6\}\}$. In this experiment, assign the number 1 to every *even* outcome, and the number 0 to every *odd* outcome. Then the **RV** $X(\zeta)$ is given by:

$$X(\zeta_1) = X(\zeta_3) = X(\zeta_5) = 0 \quad \text{and} \quad X(\zeta_2) = X(\zeta_4) = X(\zeta_6) = 1 \tag{3.1}$$

⋈

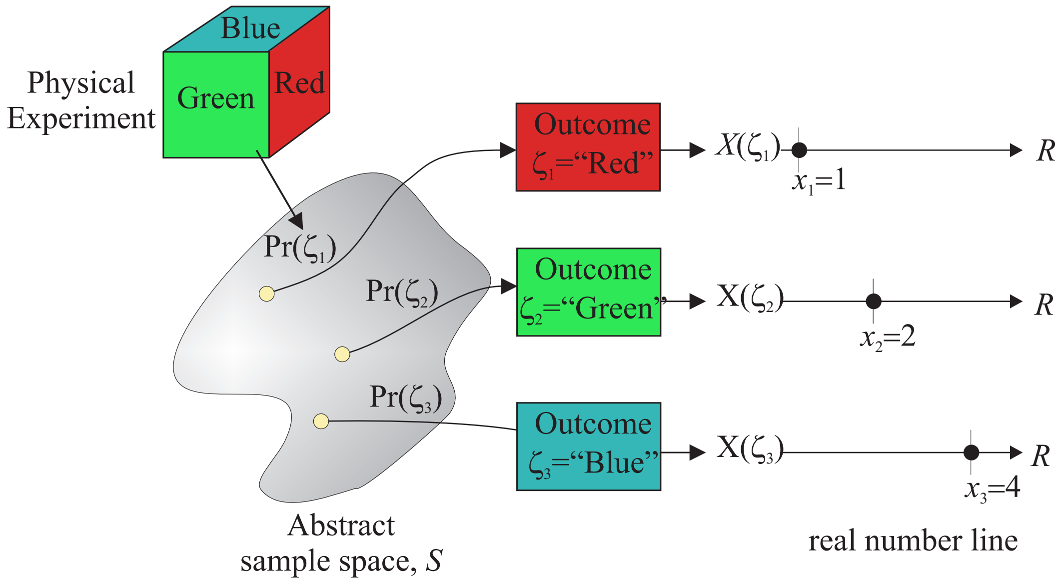


Figure 3.2: A graphical representation of a random variable.

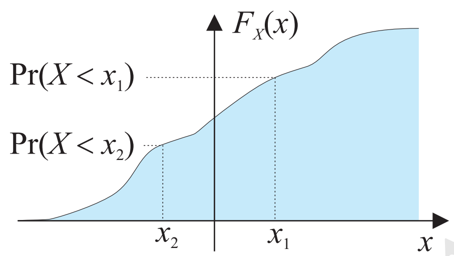


Figure 3.3: The cumulative distribution function.

Example 3.2 (Letters of the alphabet). Suppose the outcome of an experiment is a letter A to Z, such that $X(A) = 1$, $X(B) = 2$, ..., $X(Z) = 26$. Then the event $X(\zeta) \leq 5$ corresponds to the letters A, B, C, D, or E.

New slide

3.2.1 Distribution functions

Random variables are fundamentally characterised by their distribution and density functions. These concepts are considered in this and the next section.

- The **probability set function** $\Pr(X(\zeta) \leq x)$ is a function of the set $\{X(\zeta) \leq x\}$, and therefore of the point $x \in \mathbb{R}$.

- This probability is the **cumulative distribution function (cdf)**, $F_X(x)$ of a **RV** $X(\zeta)$, and is defined by:

$$F_X(x) \triangleq \Pr(X(\zeta) \leq x) \quad \text{(M:3.1.1)}$$

  It is graphically shown in Figure 3.3.

- It hence follows that the probability of being within an interval $(x_\ell, x_r]$ is given by:

$$\Pr(x_\ell < X(\zeta) \leq x_r) = \Pr(X(\zeta) \leq x_r) - \Pr(X(\zeta) \leq x_\ell) \quad (3.2)$$
$$= F_X(x_r) - F_X(x_\ell) \quad (3.3)$$

- For small intervals, it is clearly apparent that gradients are important.

This is only the first 50 pages of a 192 page handout.

Please contact the author if you would like to know more:

James R. Hopgood
james.hopgood@ed.ac.uk