Probability, Random Variables and Signals, and Classical Estimation Theory

UDRC-EURASIP Summer School, 28th June 2021

Dr James R. Hopgood

James.Hopgood@ed.ac.uk

Room 2.05

Alexander Graham Bell Building

The King's Buildings

Institute for Digital Communications

School of Engineering

College of Science and Engineering

University of Edinburgh

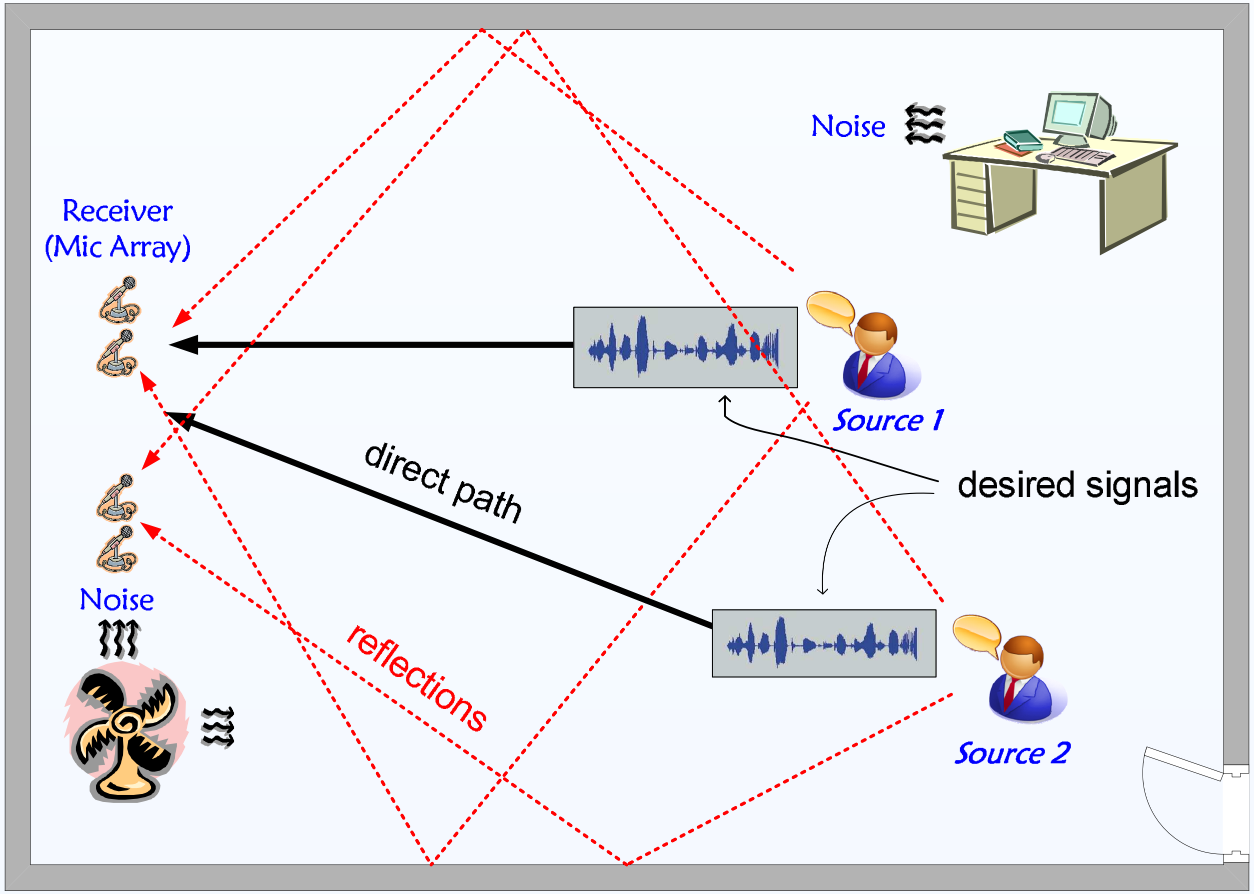
Blank Page

This slide is intentionally left blank, in order to accommodate the processing of acronyms in LaTeX.

Course overview and exemplar applications

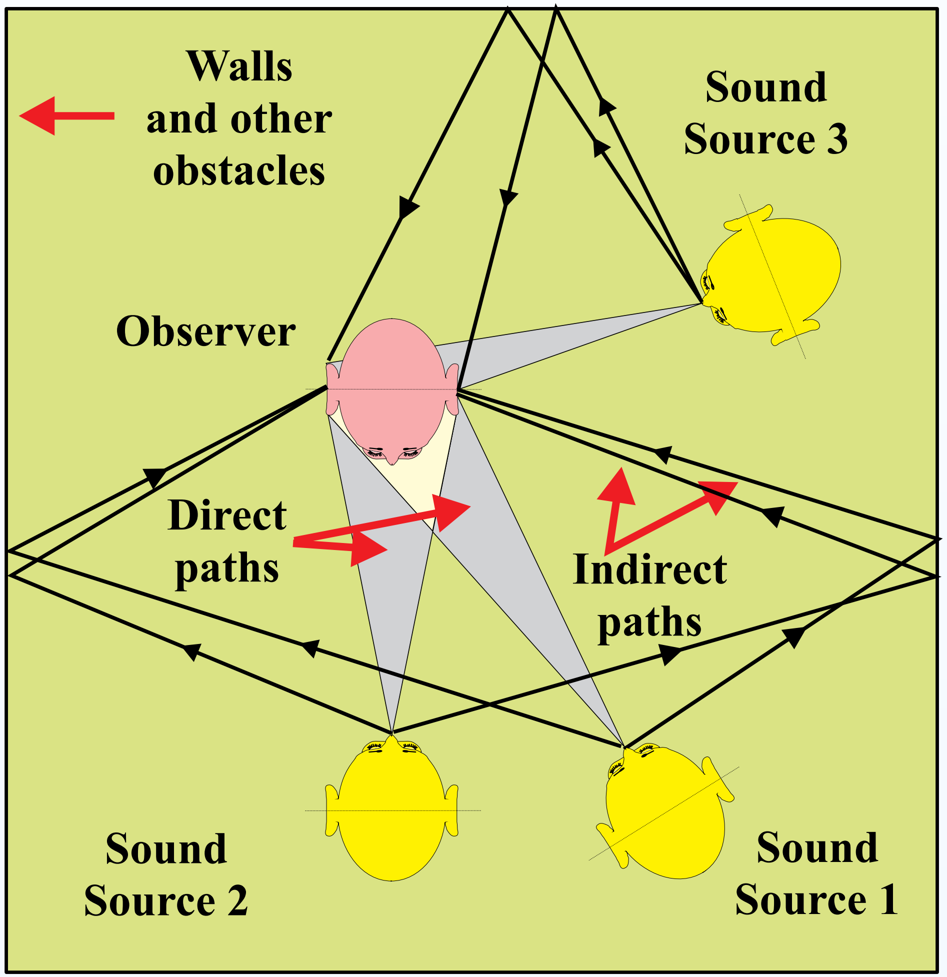
Lecture Slideset 1
Aims and Objectives

Obtaining the Latest Handouts



Source localisation and blind source separation (BSS). An example of topics using statistical signal processing.

Obtaining the Latest Handouts



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.

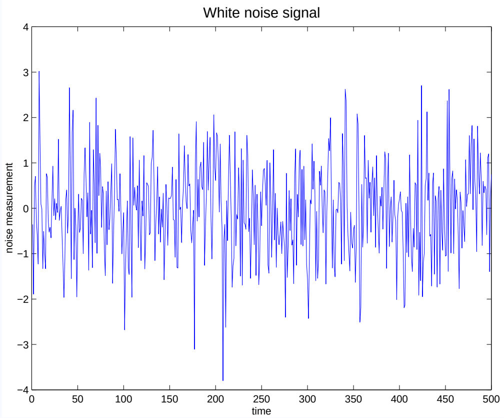
Obtaining the Latest Handouts

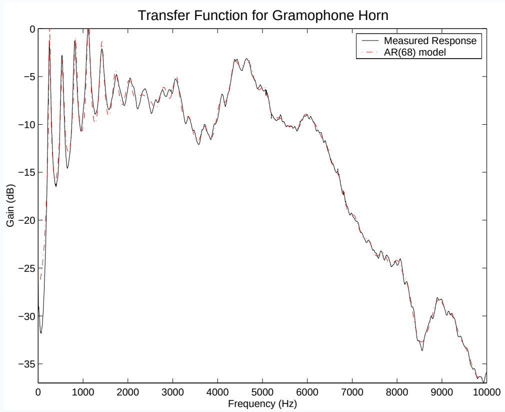
- This research tutorial is intended to cover a wide range of aspects which cover the fundamentals of statistical signal processing.
- This tutorial is being continually updated, and feedback is welcomed. The hardcopy documents published or online may differ slightly to the slides presented on the day.
- The latest version of this document can be obtained from the author, Dr James R. Hopgood, by emailing him at:

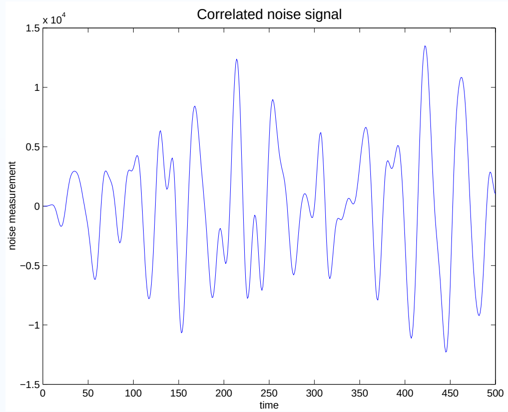
  `mailto:james.hopgood@ed.ac.uk`

  (Update: The notes are no longer online due to the desire to maintain copyright control on the document.)
- Extended thanks to the many MSc students over the past 16 years who have helped improve these documents.

Introduction and Overview







Source Signal
e.g. Clean Speech → **Channel** e.g. Room Acoustics → Observed Signal
e.g. Reverberant Speech

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.

It is assumed you have a grounding in DSP. This module will take you to the next level; a tour of the exciting, fascinating, and active research area of *statistical signal processing*.

- Course overview and exemplar applications

Aims and Objectives

- Obtaining the Latest Handouts
- Introduction and Overview
- Module Abstract
- Description and Learning Outcomes
- Structure of the Module

Signal Processing

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

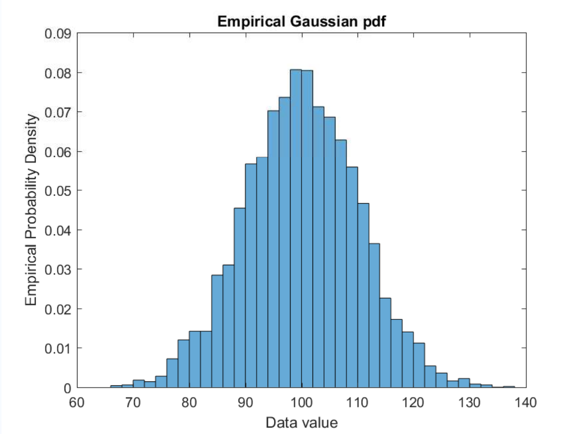
Stochastic Processes

Power Spectral Density

Linear Systems Theory

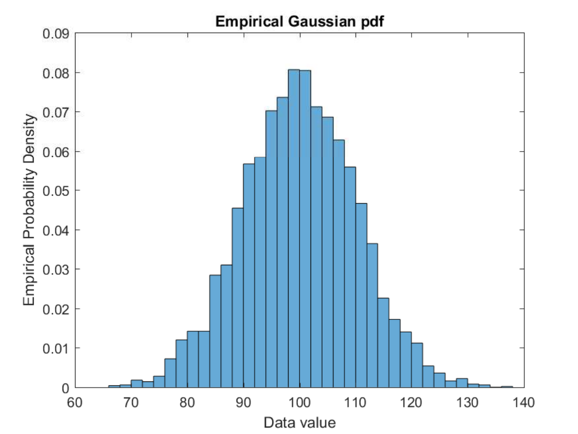
Passive Target Localisation

Module Abstract



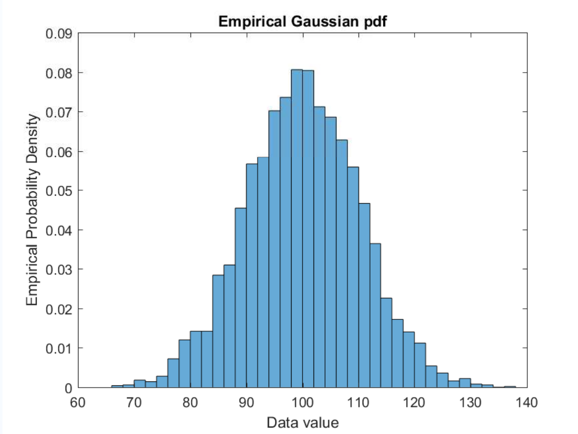
- **Random signals** are extensively used in algorithms, and are:
  - constructively used to model real-world processes;
  - described using *probability and statistics*.

Module Abstract



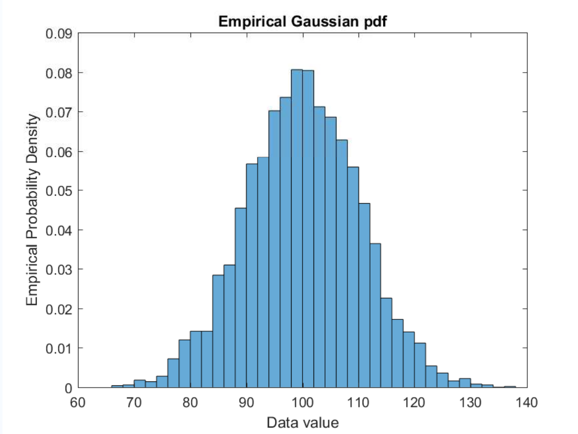
- Their properties are often estimated by assumming:
  - an infinite or large number of observations or data points;
  - time-invariant statistics.

Module Abstract



- Their properties are often estimated by assumming:
  - an infinite or large number of observations or data points;
  - time-invariant statistics.
- In practice, these statistics must be estimated from short finite-length data signals in noise.

Module Abstract



- Their properties are often estimated by assumming:
  - an infinite or large number of observations or data points;
  - time-invariant statistics.
- In practice, these statistics must be estimated from short finite-length data signals in noise.
- This module investigates relevant statistical properties, how they are estimated from real signals, and how they are used.

Description and Learning Outcomes

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

Description and Learning Outcomes

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

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;

Description and Learning Outcomes

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

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.

Description and Learning Outcomes

PETARS Learning Outcomes On completion of this course:

- Define, understand and manipulate scalar and multiple random variables, using the theory of probability; this should include the basic tools of probability transformations and characteristic functions, moments, the central limit theorem (CLT) and its use in estimation theory and the sum of random variables.

Description and Learning Outcomes

PETARS Learning Outcomes On completion of this course:

- Define, understand and manipulate scalar and multiple random variables, using the theory of probability; this should include the basic tools of probability transformations and characteristic functions, moments, the central limit theorem (CLT) and its use in estimation theory and the sum of random variables.

- Understand the principles of estimation theory, and estimation techniques such as maximum-likelihood, least squares, minimum variance unbiased estimator (MVUE) estimators, and Bayesian estimation; be able to characterise the estimator using standard metrics, including the Cramér-Rao lower-bound (CRLB).

Description and Learning Outcomes

PETARS Learning Outcomes On completion of this course:

- Explain, describe, and understand the notion of a random process and statistical time series, and characterise them in terms of its statistical properties.

Description and Learning Outcomes

PETARS Learning Outcomes On completion of this course:

- Explain, describe, and understand the notion of a random process and statistical time series, and characterise them in terms of its statistical properties.

- Define, describe, and understand the notion of the power spectral density of stationary random processes, and be able to analyse and manipulate them; 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.

Description and Learning Outcomes

PETARS Learning Outcomes On completion of this course:

- 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; apply least squares, maximum-likelihood, and Bayesian estimators to model based signal processing problems.

Structure of the Module

The key **themes** covered are:

1. review of the fundamentals of **probability theory**;

Structure of the Module

The key **themes** covered are:

1. review of the fundamentals of **probability theory**;
2. **random variables** and **stochastic processes**;

Structure of the Module

The key **themes** covered are:

1. review of the fundamentals of **probability theory**;
2. **random variables** and **stochastic processes**;
3. principles of **estimation theory**;

Structure of the Module

The key **themes** covered are:

1. review of the fundamentals of **probability theory**;
2. **random variables** and **stochastic processes**;
3. principles of **estimation theory**;
4. **Bayesian estimation** theory;

Structure of the Module

The key **themes** covered are:

1. review of the fundamentals of **probability theory**;
2. **random variables** and **stochastic processes**;
3. principles of **estimation theory**;
4. **Bayesian estimation** theory;
5. review of **Fourier transforms** and **discrete-time systems**;

Structure of the Module

The key **themes** covered are:

1. review of the fundamentals of **probability theory**;
2. **random variables** and **stochastic processes**;
3. principles of **estimation theory**;
4. **Bayesian estimation** theory;
5. review of **Fourier transforms** and **discrete-time systems**;
6. **linear systems** with stationary random inputs, and **linear system models**;

Structure of the Module

The key **themes** covered are:

1. review of the fundamentals of **probability theory**;
2. **random variables** and **stochastic processes**;
3. principles of **estimation theory**;
4. **Bayesian estimation** theory;
5. review of **Fourier transforms** and **discrete-time systems**;
6. **linear systems** with stationary random inputs, and **linear system models**;
7. **signal modelling** and **parametric spectral estimation**;

Structure of the Module

The key **themes** covered are:

1. review of the fundamentals of **probability theory**;
2. **random variables** and **stochastic processes**;
3. principles of **estimation theory**;
4. **Bayesian estimation** theory;
5. review of **Fourier transforms** and **discrete-time systems**;
6. **linear systems** with stationary random inputs, and **linear system models**;
7. **signal modelling** and **parametric spectral estimation**;
8. an application investigating the estimation of sinusoids in noise, outperforming the Fourier transform.

Structure of the Module

– End-of-Topic 1: **Course description, learning outcomes, and prerequisites** –

Any Questions?

Lecture Slideset 2
Signal Processing

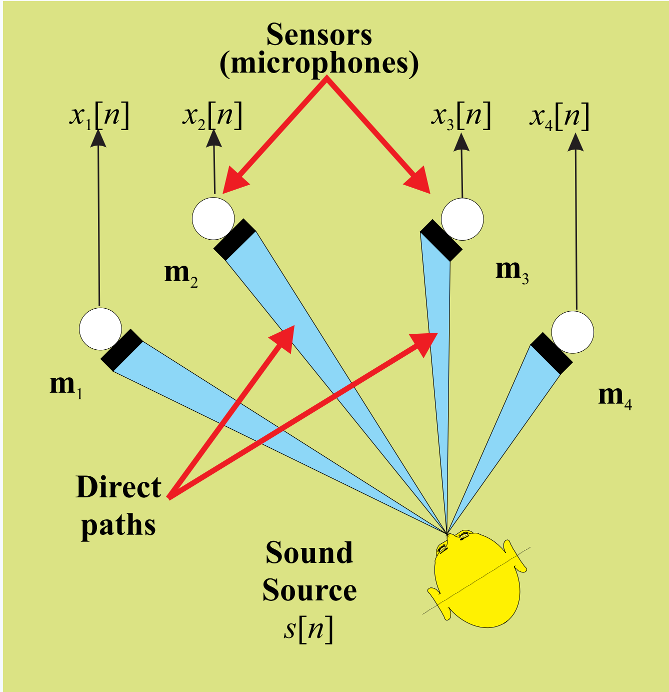
Passive and Active Target Localisation

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

1. Tracking methods and target intent inference.
2. Estimating 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.

- Passive localisation is particularly challenging.

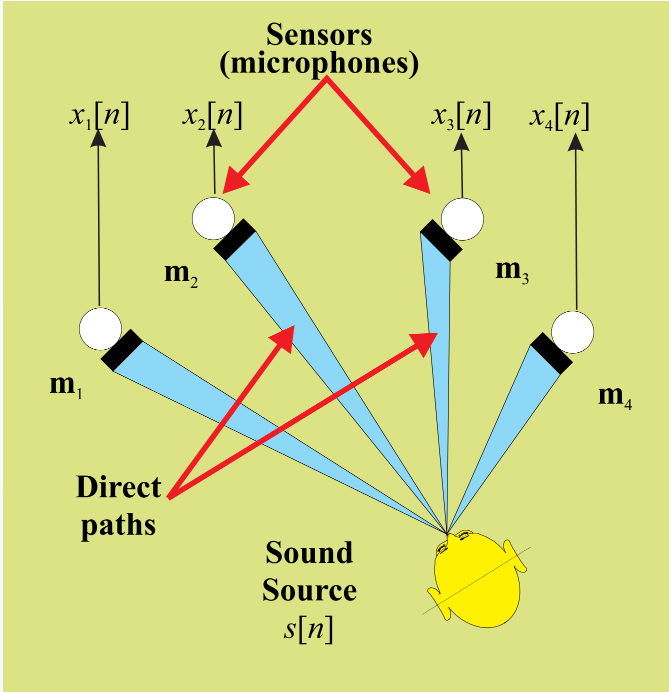
Passive Target Localisation Methodology



Ideal free-field model.

- Most passive target localisation (PTL) techniques rely on the fact that an impinging wavefront reaches one sensor before it reaches another (spatio-temporal diversity).

Passive Target Localisation Methodology



Ideal free-field model.

- Most PTL techniques rely on the fact that an impinging wavefront reaches one sensor before it reaches another (spatio-temporal diversity).
- Many PTL algorithms are designed assuming there is no multipath or reverberation present, the *free-field assumption*.

Source Localization Strategies

Existing source localisation methods can loosely be divided into:

1. those based on maximising the steered response power (SRP) of a beamformer:
   - location estimate derived directly from a filtered, weighted, and summed version of the signal data;

Source Localization Strategies

Existing source localisation methods can loosely be divided into:

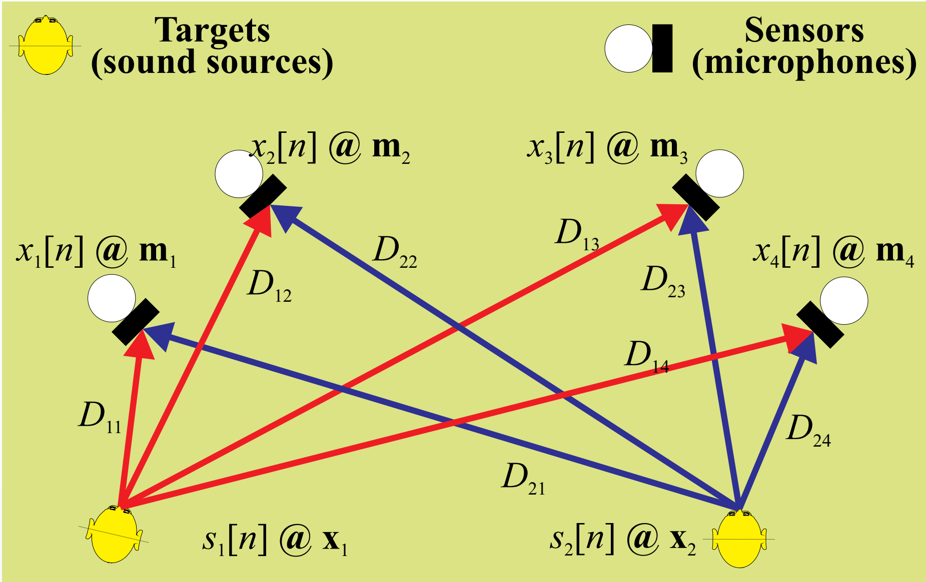
1. those based on maximising the steered response power (SRP) of a beamformer:
   - location estimate derived directly from a filtered, weighted, and summed version of the signal data;
2. techniques adopting high-resolution spectral estimation concepts:
   - any localisation scheme relying upon an application of the signal correlation matrix;

Source Localization Strategies

Existing source localisation methods can loosely be divided into:

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

Geometric Layout

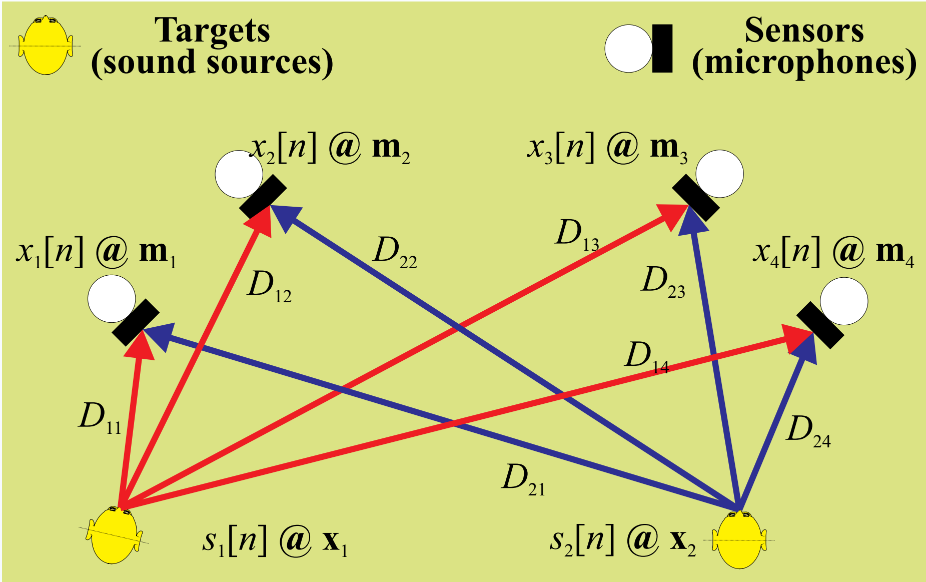


Geometry assuming a free-field model.

Suppose there is a:

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

Geometric Layout



Geometry assuming a free-field model.

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:

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

where c is the speed of the impinging wavefront.

Ideal Free-field Model

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

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

  where $b_{ik}(t)$ denotes additive noise, and α_{ik} is the attenuation.

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

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

- The additive noise source is assumed to be uncorrelated with the source and noise sources at other sensors.

Ideal Free-field Model

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

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

  where $b_{ik}(t)$ denotes additive noise, and α_{ik} is the attenuation.

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

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

- 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:

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

Indirect TDOA-based Methods

This is typically a two-step procedure in which:

- Typically, TDOAs are extracted using the GCC function, or an adaptive eigenvalue decomposition (AED) algorithm.

Indirect TDOA-based Methods

This is typically a two-step procedure in which:

- Typically, TDOAs are extracted using the 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.

Indirect TDOA-based Methods

This is typically a two-step procedure in which:

- Typically, TDOAs are extracted using the 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.

Indirect TDOA-based Methods

This is typically a two-step procedure in which:

- Typically, TDOAs are extracted using the 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.

Indirect TDOA-based Methods

This is typically a two-step procedure in which:

- Typically, TDOAs are extracted using the 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.

Hyperbolic Least Squares Error Function

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

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

- The total error as a function of target position

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

where

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

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

TDOA estimation methods

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

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

TDOA estimation methods

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

- 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.

TDOA estimation methods

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

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.

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

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

- The cross-correlation function is given by

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

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

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

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

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

- The cross-correlation function is given by

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

where the CPSD is given by

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

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

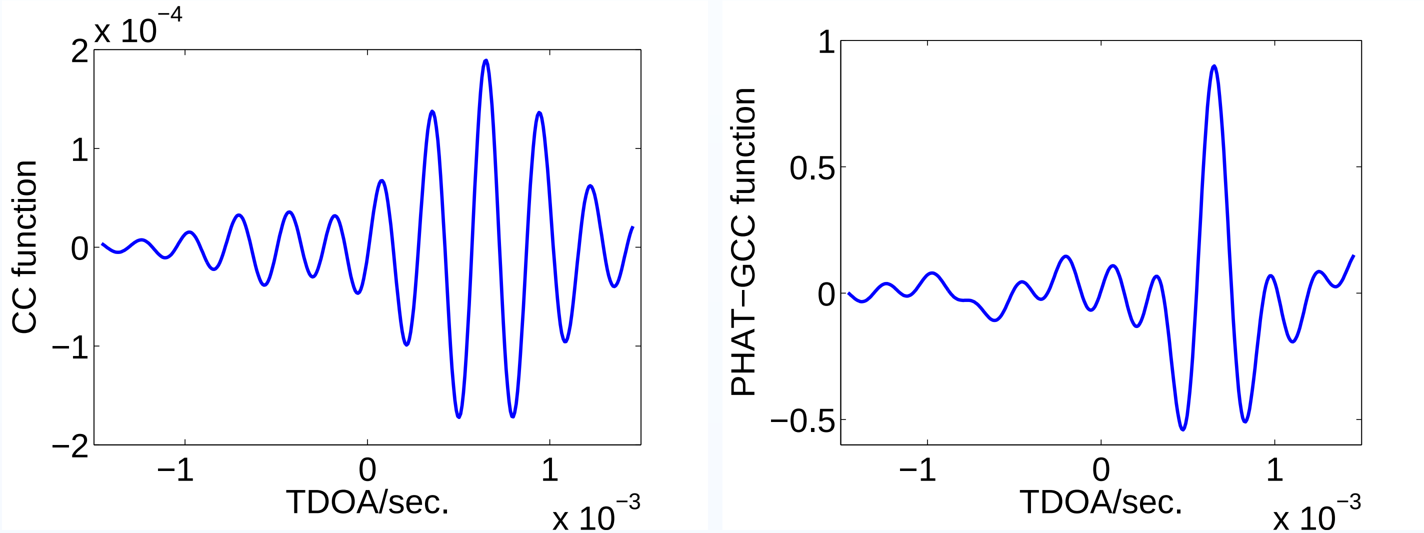
$$\angle P_{x_i x_j}(\omega) = -j\omega\, T\left(\mathbf{m}_i,\, \mathbf{m}_j,\, \mathbf{x}_k\right)$$

GCC Processors

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

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

GCC Processors



Normal cross-correlation and GCC-phase transform (PHAT) (GCC-PHAT) functions for a frame of speech.

Direct Localisation Methods

- 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.

Steered Response Power Function

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 $\hat{\tau}_{pk} = |\hat{\mathbf{x}} - \mathbf{m}_p|$:

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

Steered Response Power Function

The 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.

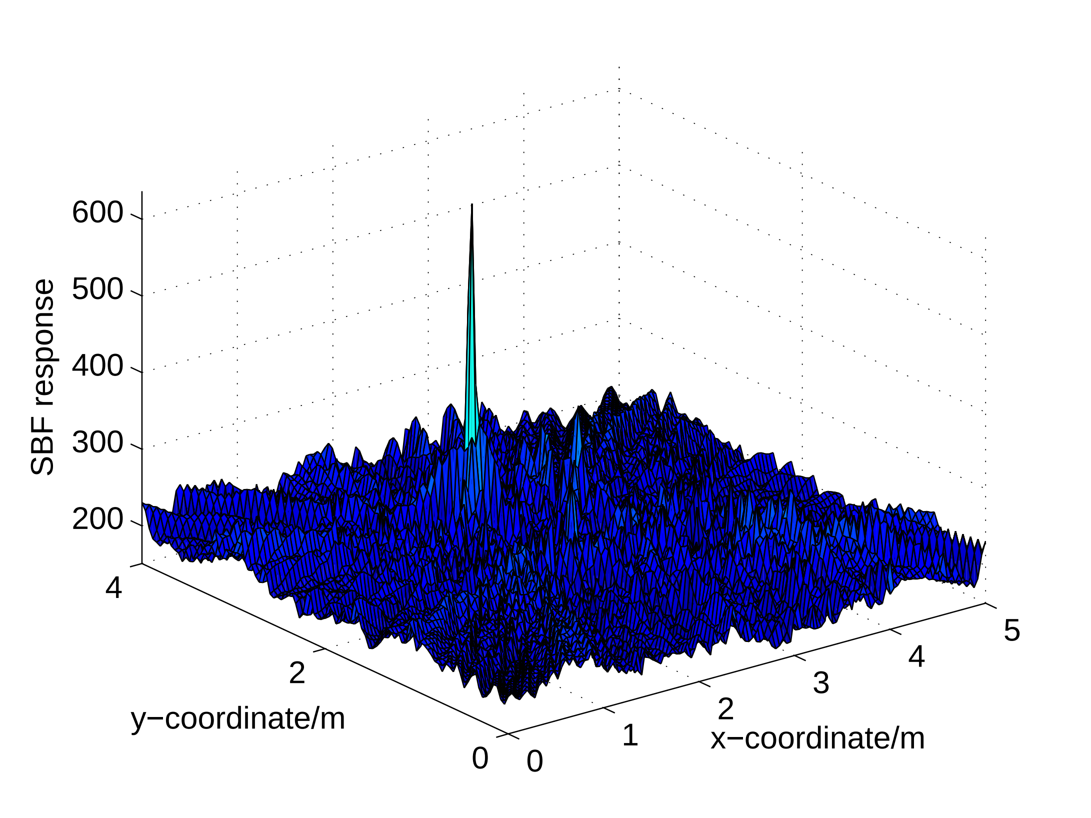
The frequency domain **delay-and-sum beamformer** steered to a spatial position $\hat{\mathbf{x}}_k$ such that $\hat{\tau}_{pk} = |\hat{\mathbf{x}} - \mathbf{m}_p|$:

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

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

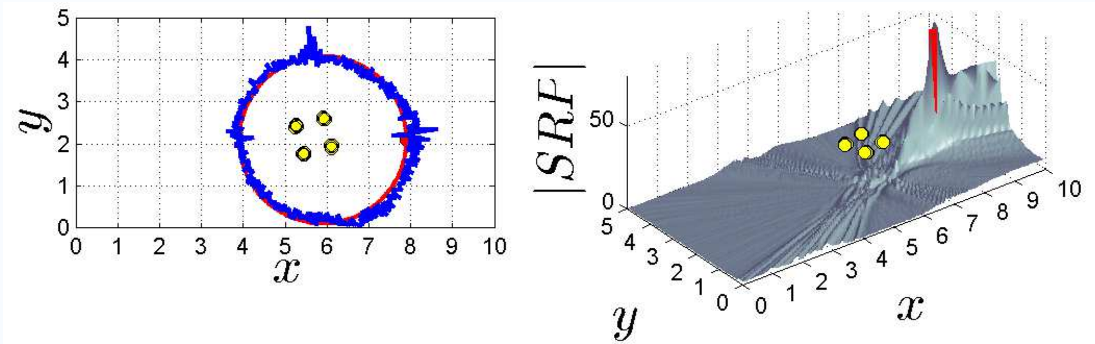
$$\equiv \sum_{p=1}^{N}\sum_{q=1}^{N} r_{x_i\,x_j}\left[\frac{|\mathbf{x}_k - \mathbf{m}_i| - |\mathbf{x}_k - \mathbf{m}_j|}{c}\right]$$

Steered Response Power Function



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

Steered Response Power Function



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

- Show video!

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!

Probability, Random Variables, and Estimation Theory

Lecture Slideset 1
Probability Theory

Introduction

To motivate the need for probability theory, consider the simplest of problems in the presence of uncertainty.

How many water taxis are there in Venice?

Introduction

To motivate the need for probability theory, consider the simplest of problems in the presence of uncertainty.


How many water taxis are there in Venice?

Assume taxi numbers sequential from 1 to N. What is best guess of N given these observations?

Introduction

To motivate the need for probability theory, consider the simplest of problems in the presence of uncertainty.

How many water taxis are there in Venice?

Assume taxi numbers sequential from 1 to N. What is best guess of N given these observations?

Introduction

To motivate the need for probability theory, consider the simplest of problems in the presence of uncertainty.


How many water taxis are there in Venice?


How does your answer change when you see more taxis?

Introduction

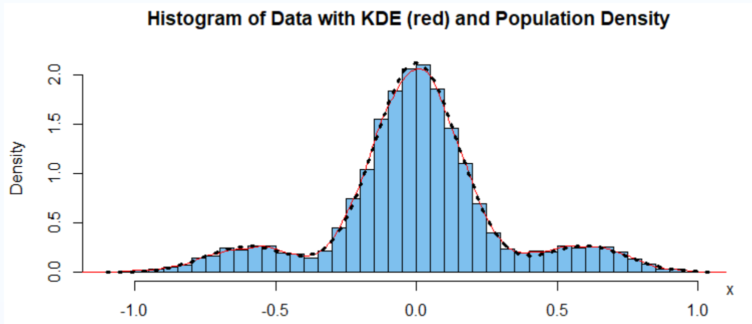
To motivate the need for probability theory, consider the simplest of problems in the presence of uncertainty.

How many water taxis are there in Venice?

How does your answer change when you see more taxis?

Introduction

What tools are needed to study this problem?

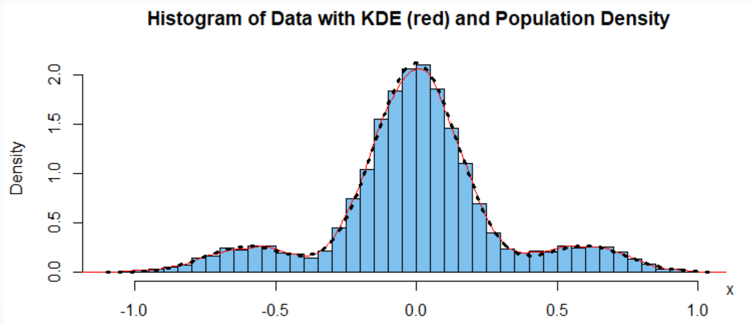


Kernel density estimation for modelling observation data.

- The notion of probability and random variables;
- The notion of probability density functions (pdfs);

Introduction

What tools are needed to study this problem?

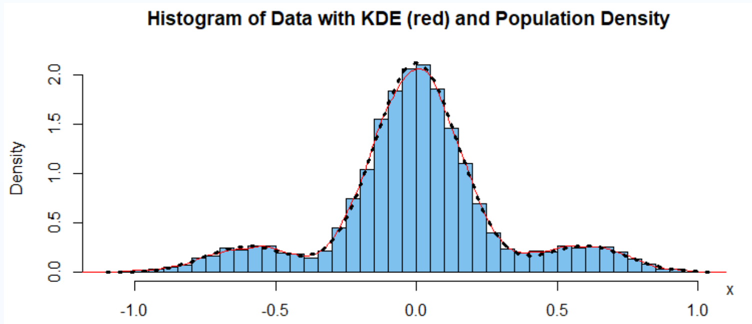


Kernel density estimation for modelling observation data.

- The notion of probability and random variables;
- The notion of probability density functions (pdfs);
- The notion of independence of observations;

Introduction

What tools are needed to study this problem?



Kernel density estimation for modelling observation data.

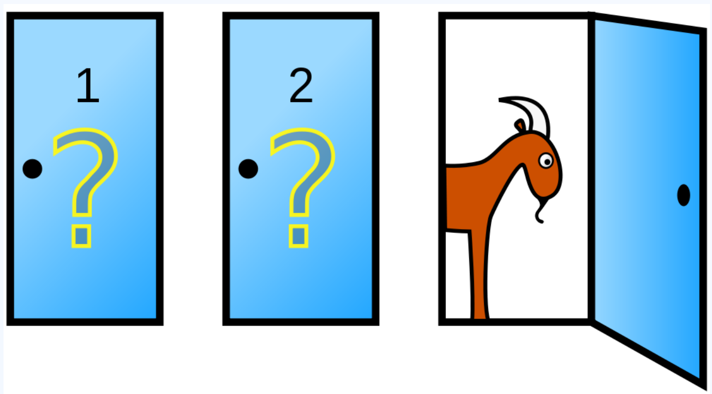
- The notion of probability and random variables;
- The notion of probability density functions (pdfs);
- The notion of independence of observations;
- The notion of estimation theory & uncertainty quantification.

These will be studies in turn throughout this course; we will start off looking at the basics of probability.

Introduction

Students are exposed to probability at school from a relatively young age. It is not the intention of this course to go over basic probability again. Instead, the purpose is to:

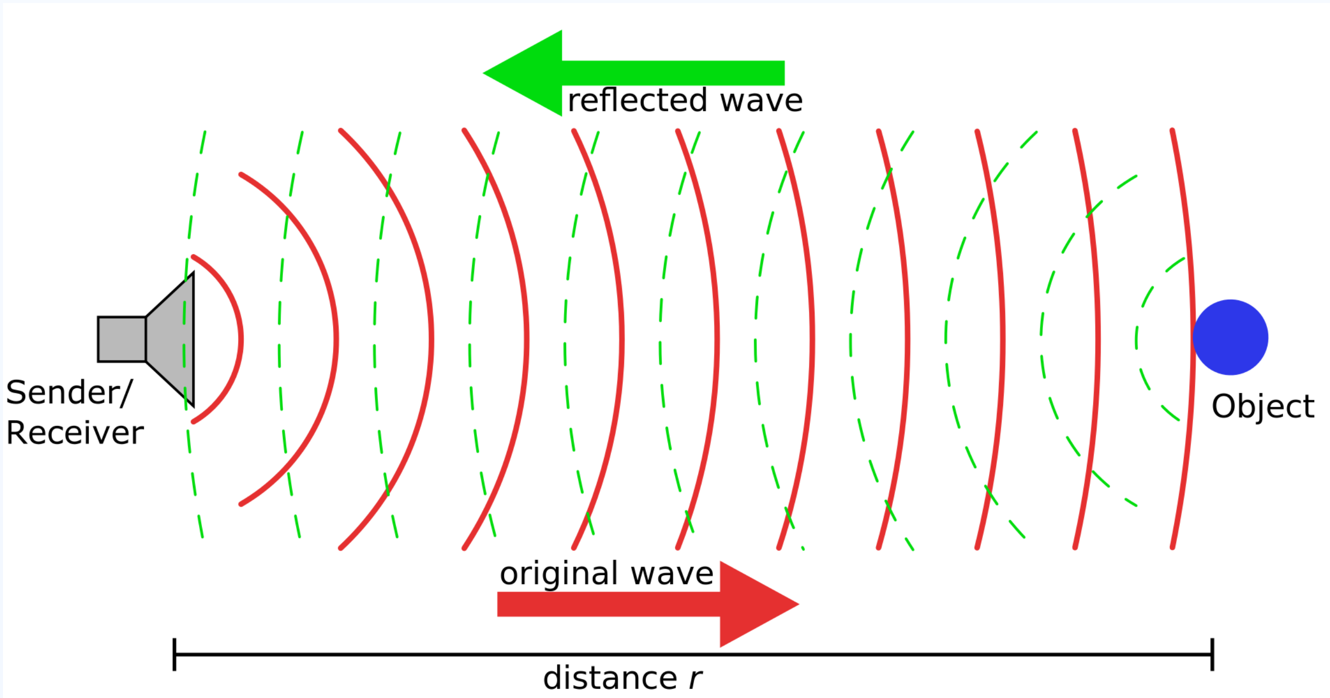
- enhance a fundamental understanding of probability that enable us develop more complex concepts;
- identify limitations of classical definitions;
- reaffirm that intuition with regards to probability is often wrong; careful and systematic analysis is often needed.



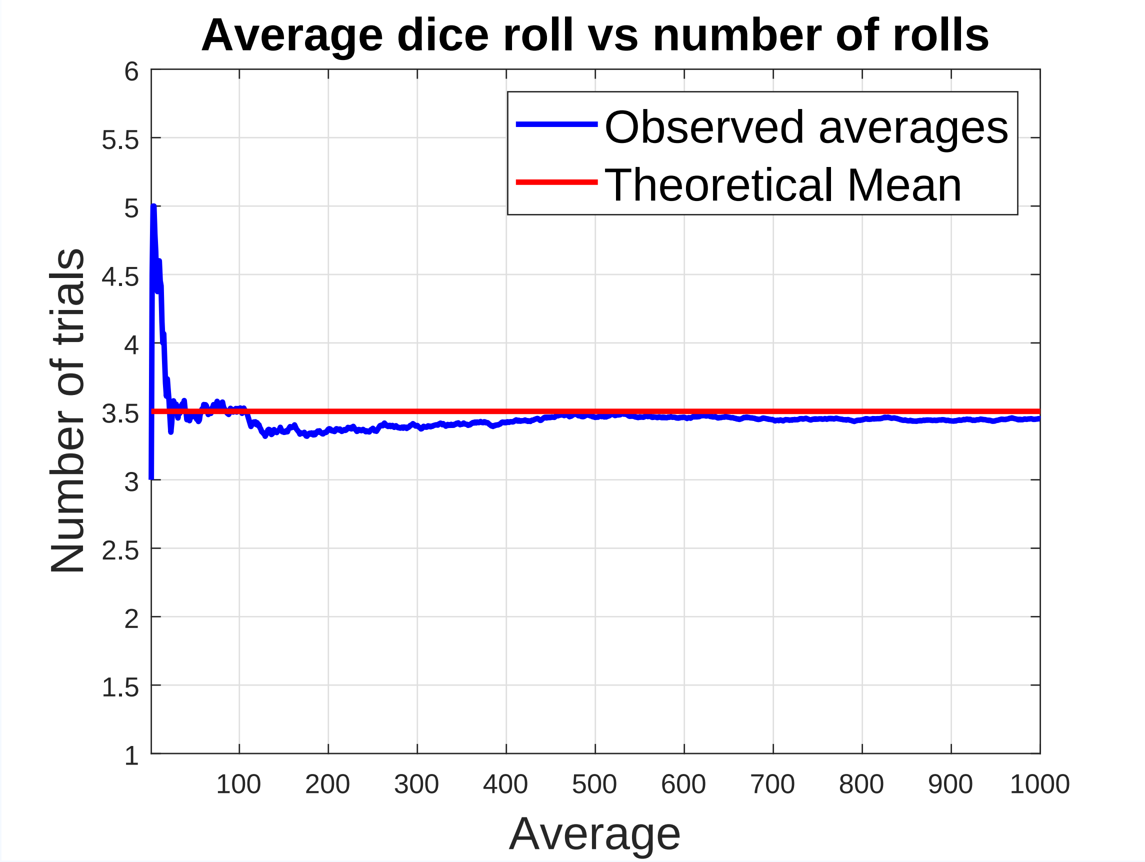
Is the infamous Monty-Hall problem counter-intuitive?

Introduction

- The theory of probability deals with averages of mass phenomena occurring sequentially or simultaneously;
  - e.g. signal/anomaly detection, parameter estimation, ...
- Starting from probability of individual events, can develop a probabilistic framework for analysing signals.



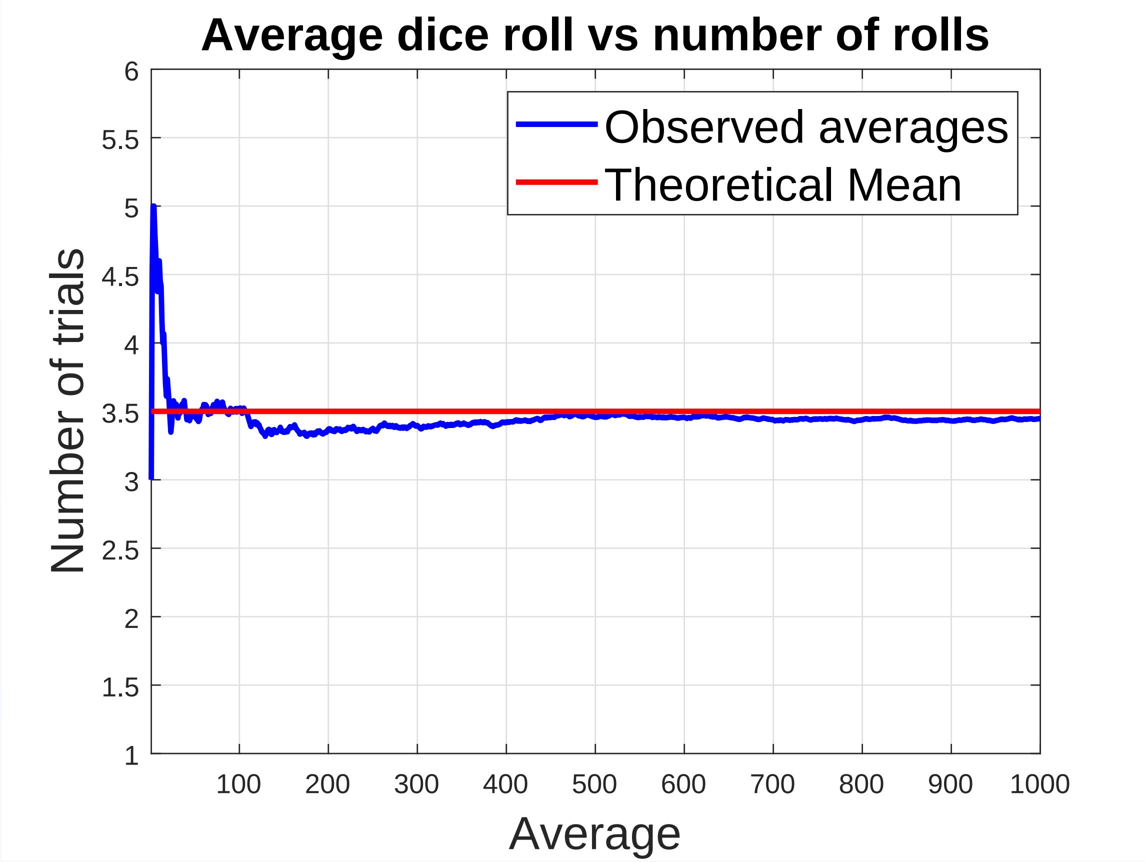
The Notion of Probability



Illustrating law-of-large numbers through throwing dice.

- Start by *observing* certain averages approach a constant value as the number of observations increases; and remains constant even if evaluated over any specified sub-sequences.

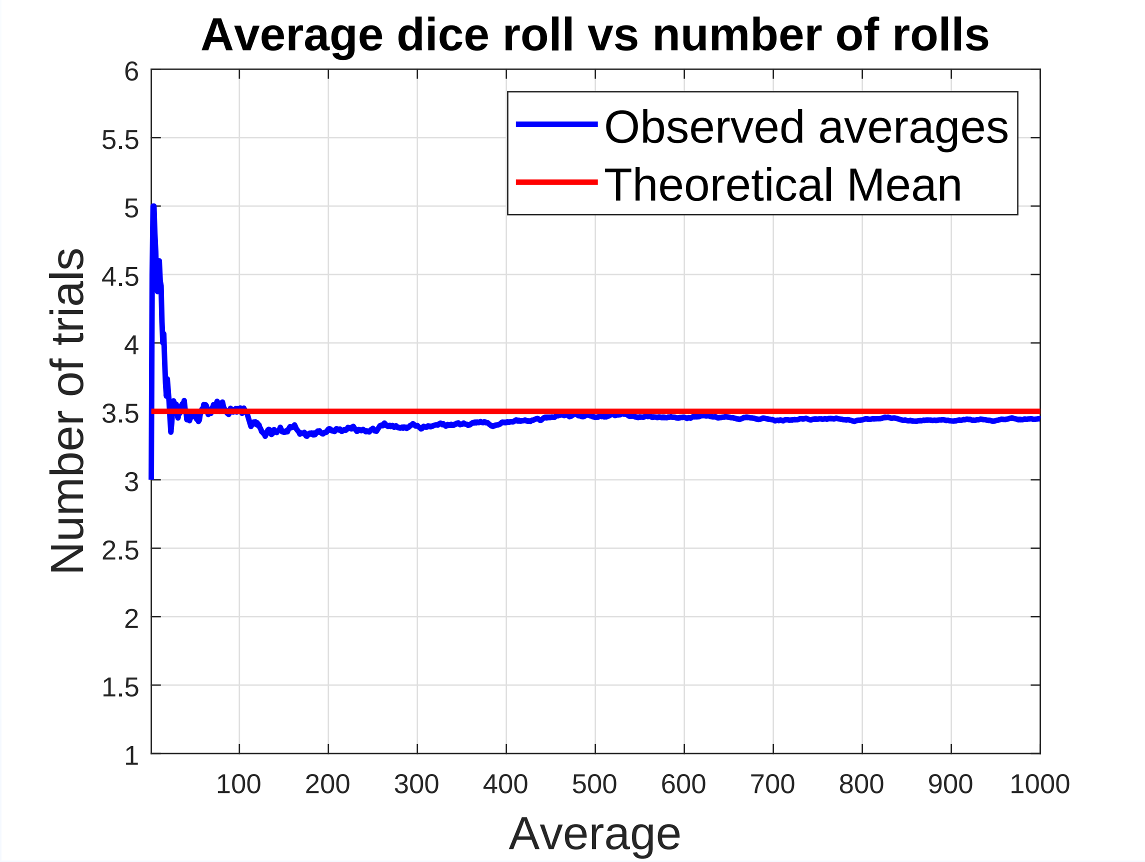
The Notion of Probability



Illustrating law-of-large numbers through throwing dice.

As the number of rolls in the sequence increases, the average of the values of all the results approaches the theoretical **mean value** of $\frac{1}{6}\sum_{k=1}^{6} k = 3.5$.

The Notion of Probability



Illustrating law-of-large numbers through throwing dice.

It follows from the law of large numbers that the **empirical probability** of success in a series of Bernoulli trials will converge to the theoretical probability.

The Notion of Probability

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:

$$\Pr(A) \approx \frac{n_A}{n}$$

provided that n is *sufficiently large*.

This is the **empirical probability**, or **relative frequency**, and is an *estimator of probability*.

Note this frequentist interpretation and language is imprecise.

The Notion of Probability

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:

$$\Pr(A) \approx \frac{n_A}{n}$$

provided that n is *sufficiently large*.

This is the **empirical probability**, or **relative frequency**, and is an *estimator of probability*.

Note this frequentist interpretation and language is imprecise.

- Moreover, another problem with this definition is that it implies an experiment needs to be performed in order to define a probability. In the next set of slides, we will move away from this restriction.

The Notion of Probability

– End-of-Topic 11: **Introduction to Probability, The Law-of-Large Numbers, and Empirical Probability** –

Any Questions?

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 determined *a priori* without actual experimentation. It is given by the ratio:

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

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*.

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 determined *a priori* without actual experimentation. It is given by the ratio:

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

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*.

1. Probability of a specific number rolled on a six-sided die ($^1/_6$);
2. Probability of rolling an even number on a six-sided die ($^3/_6$).

Difficulties with the Classical Definition

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

Difficulties with the Classical Definition

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.

Difficulties with the Classical Definition

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.

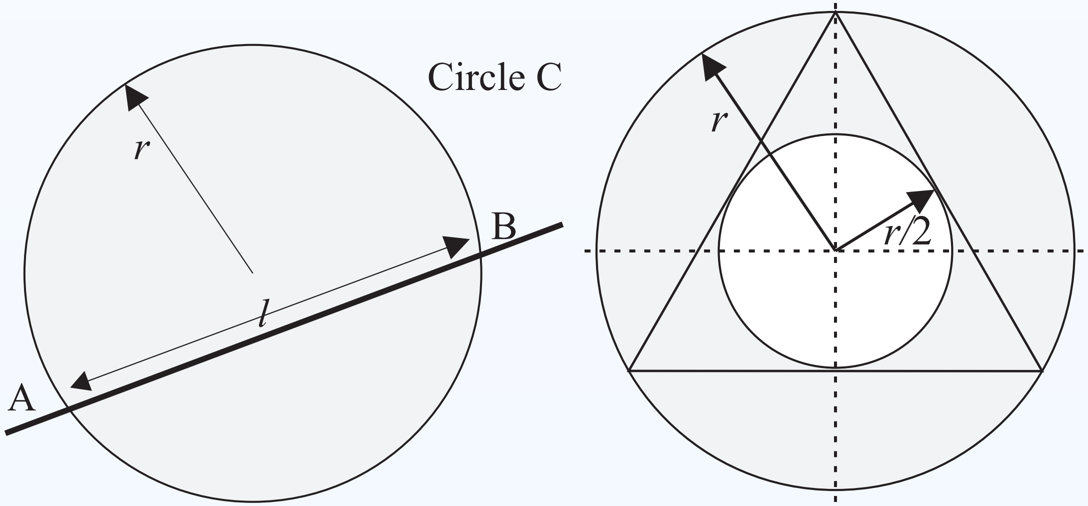
   In the die experiment, for example, it is applicable only if the six faces have the same probability. If the die is loaded and the probability of a "4" equals 0.2, say, then this cannot be determined from the classical ratio.

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

Discussion: Bertrand's Paradox

The Bertrand paradox is a problem within the classical interpretation of probability theory.

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?

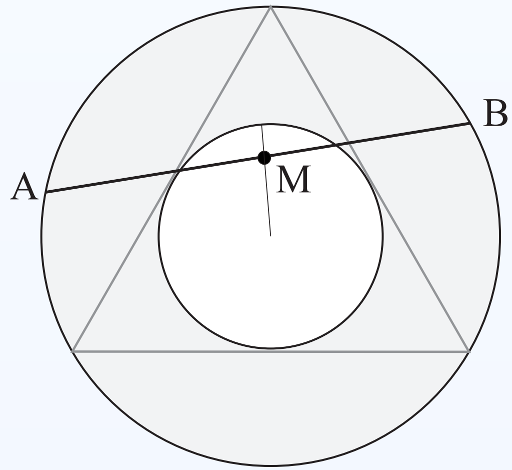


Bertrand's paradox, problem definition.

Discussion: Bertrand's Paradox

- 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, such that the resulting chord AB through these chosen points has M as its midpoint.

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

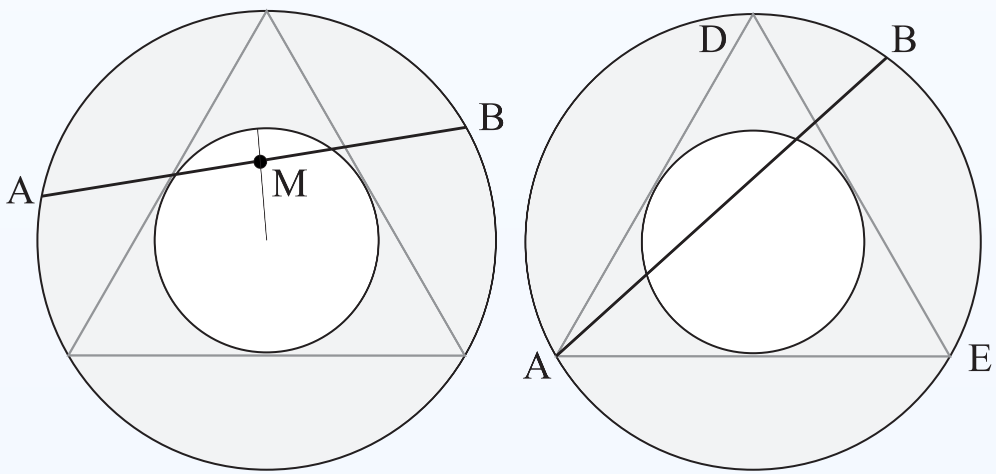


Different selection methods.

Discussion: Bertrand's Paradox

- 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.

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

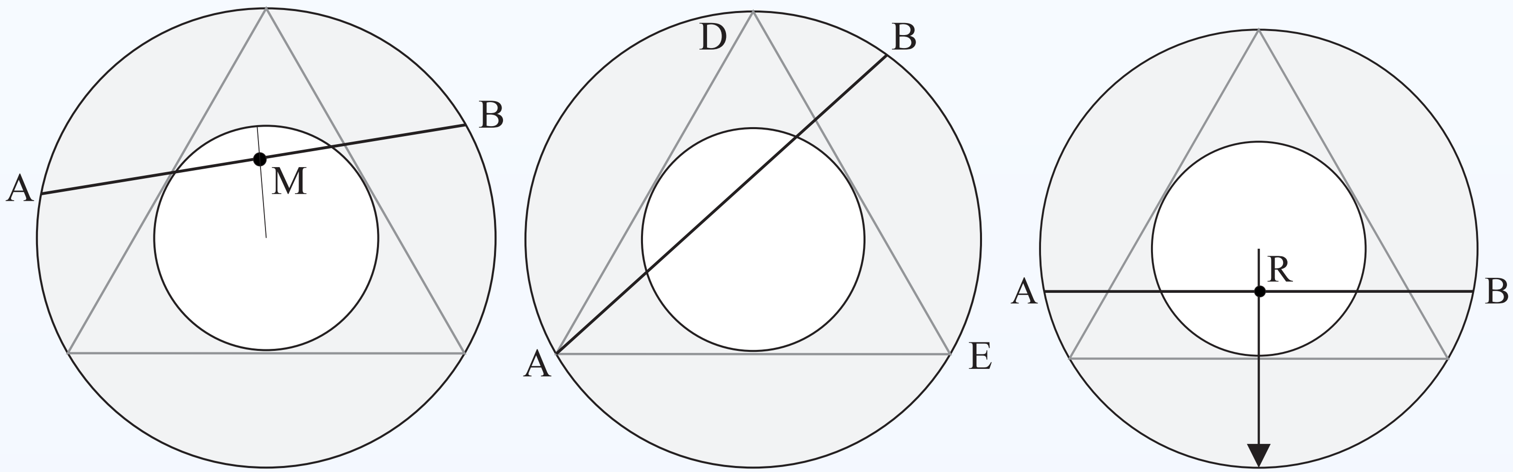


Different selection methods.

Discussion: Bertrand's Paradox

- 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.

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



Different selection methods.

There are three different reasonable solutions. Which is valid?

Discussion: Bertrand's Paradox

Example (Multi-choice). Consider a circle of radius r. What is the probability that the length of a *randomly selected* cord is greater than the length, $r\sqrt{3}$, of the inscribed equilateral triangle?

1. $\frac{1}{4}$
2. $\frac{1}{3}$
3. $\frac{1}{2}$
4. Need more information.

Discussion: Bertrand's Paradox

Example (Multi-choice). Consider a circle of radius r. What is the probability that the length of a *randomly selected* cord is greater than the length, $r\sqrt{3}$, of the inscribed equilateral triangle?

1. $\frac{1}{4}$
2. $\frac{1}{3}$
3. $\frac{1}{2}$
4. Need more information.

The solution to this paradox is indeed quite complicated, and has been discussed in a number of research papers! A discussion will take place in the hybrid classes, but if you are interested in finding out more, you are encouraged to look into this further.

Discussion: Bertrand's Paradox

– End-of-Topic 12: **Awareness of the difficulties with the Classical Definition of Probability** –

Any Questions?

Axiomatic Definition

The Kolmogorov axioms are the foundations of probability introduced in 1933. An alternative approach is Cox's theorem.

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

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

$$\Pr\left(A\right) \geq 0$$

Axiomatic Definition

The Kolmogorov axioms are the foundations of probability introduced in 1933. An alternative approach is Cox's theorem.

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:

$$\Pr(A) \geq 0$$

2. Defining the **certain event**, S, as the event that occurs in every trial, then:

$$\Pr(S) = 1$$

Axiomatic Definition

The Kolmogorov axioms are the foundations of probability introduced in 1933. An alternative approach is Cox's theorem.

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:

$$\Pr(A) \geq 0$$

2. Defining the **certain event**, S, as the event that occurs in every trial, then:

$$\Pr(S) = 1$$

3. If the events A and B are **mutually exclusive**, then:

$$\Pr(A \cup B) = \Pr(A) + \Pr(B)$$

Properties of Axiomatic Probability

Impossible Event The probability of the impossible event is 0:

$$\Pr(\emptyset) = 0$$

Properties of Axiomatic Probability

Impossible Event The probability of the impossible event is 0:

$$\Pr(\emptyset) = 0$$

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

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

Properties of Axiomatic Probability

Impossible Event The probability of the impossible event is 0:

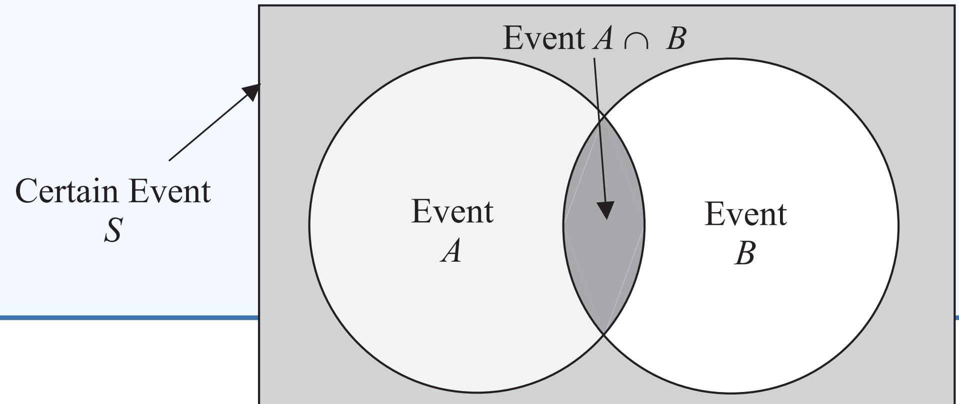
$$\Pr(\emptyset) = 0$$

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

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

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

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



Properties of Axiomatic Probability

Example (Sum Rule). Let A and B be events with probabilities $\Pr(A) = {}^3\!/_4$ and $\Pr(B) = {}^1\!/_3$. Show that ${}^1\!/_{12} \leq \Pr(A\,B) \leq {}^1\!/_3$.

Properties of Axiomatic Probability

Example (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:

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

which is the case when the whole **sample space** is covered by the two events.

Properties of Axiomatic Probability

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

SOLUTION. Using the sum rule, that:

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

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 $\Pr(A\,B) \leq \min\{\Pr(A), \Pr(B)\} = 1/3$. □

Properties of Axiomatic Probability

– End-of-Topic 13: **Properties of axiomatic probability theory, and an interesting example** –

Any Questions?

Set Theory

Unions & Intersections Unions and intersections are commutative, associative, distributive:

$$A \cup B = B \cup A, \quad (A \cup B) \cup C = A \cup (B \cup C)$$

$$AB = BA, \quad (AB)C = A(BC), \quad A(B \cup C) = AB \cup AC$$

Set Theory

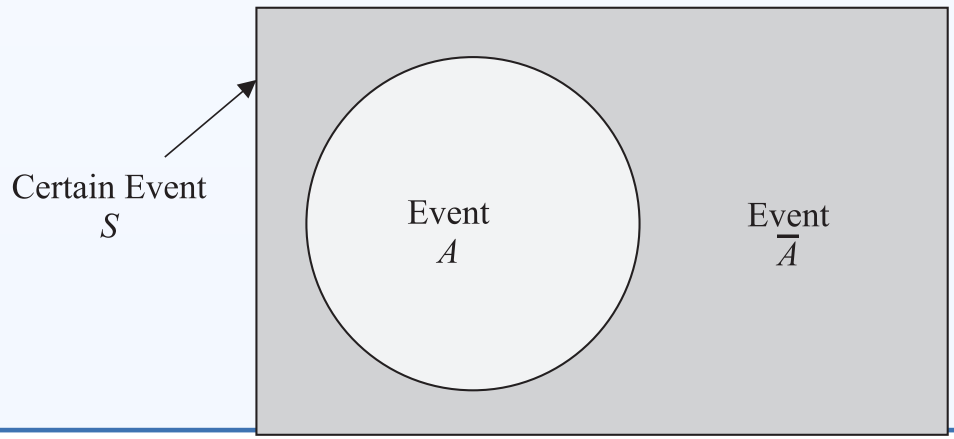
Unions & Intersections Unions and intersections are commutative, associative, distributive:

$$A \cup B = B \cup A, \quad (A \cup B) \cup C = A \cup (B \cup C)$$

$$AB = BA, \quad (AB)C = A(BC), \quad A(B \cup C) = AB \cup AC$$

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:

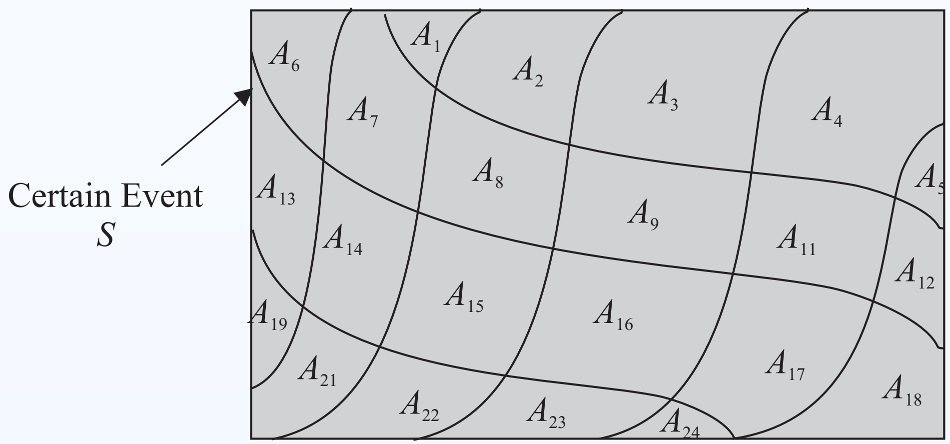
$$A \cup \overline{A} = S \quad \text{and} \quad A \cap \overline{A} \equiv A\overline{A} = \{\emptyset\}$$



Set Theory

Partitions A partition U of a set S is a collection of mutually exclusive subsets A_i of S whose union equates to S:

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



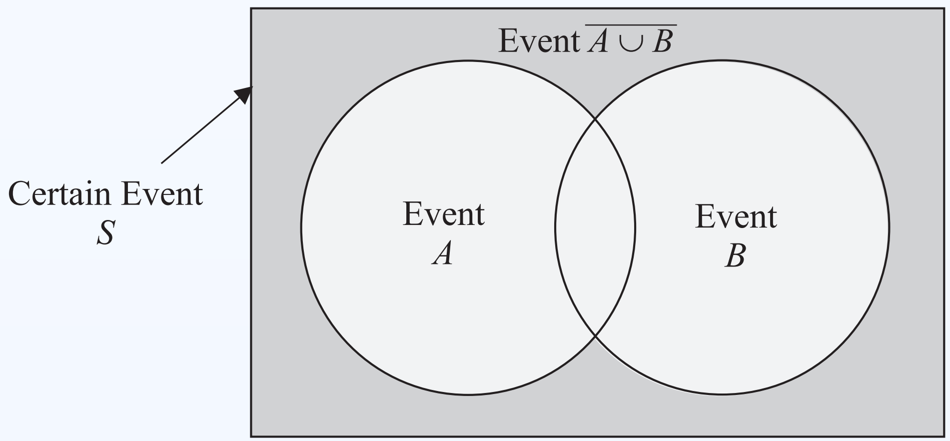
Set Theory

Partitions A partition U of a set S is a collection of mutually exclusive subsets A_i of S whose union equates to S:

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

De Morgan's Law Using Venn diagrams, it it can be shown

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



Set Theory

Partitions A partition U of a set S is a collection of mutually exclusive subsets A_i of S whose union equates to S:

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

De Morgan's Law Using Venn diagrams, it it can be shown

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

As an application of this, note that:

$$\begin{aligned} \overline{A \cup BC} &= \overline{A}\,\overline{BC} = \overline{A}\left(\overline{B} \cup \overline{C}\right) \\ &= \left(\overline{A}\,\overline{B}\right) \cup \left(\overline{A}\,\overline{C}\right) = \overline{A \cup B} \cup \overline{A \cup C} \\ \Rightarrow \quad A \cup BC &= (A \cup B)\,(A \cup C) \end{aligned}$$

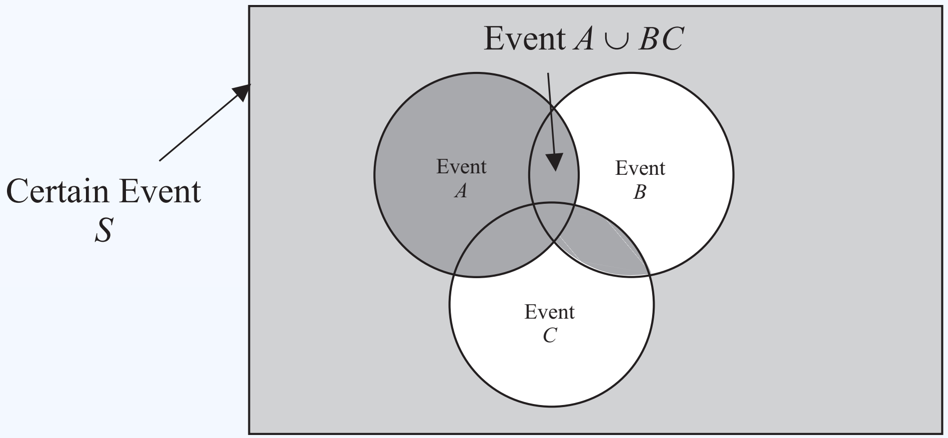
Set Theory

De Morgan's Law Using Venn diagrams, it it can be shown

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

As an application of this, note that:

$$\begin{aligned} \overline{A \cup BC} &= \overline{A}\,\overline{BC} = \overline{A}\left(\overline{B} \cup \overline{C}\right) \\ &= \left(\overline{A}\,\overline{B}\right) \cup \left(\overline{A}\,\overline{C}\right) = \overline{A \cup B} \cup \overline{A \cup C} \\ \Rightarrow \quad A \cup BC &= (A \cup B)\,(A \cup C) \end{aligned}$$



Set Theory

Example (Proof of the Sum Rule). Prove the addition law of probability (or sum rule), namely:

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

SOLUTION.

Set Theory

Example (Proof of the Sum Rule). SOLUTION. To prove this, separately write *each of* $A \cup B$ and B as the union of two mutually exclusive events.

- First, to write $A \cup B$ in this way, use S:

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

Since the intersection $A \cap \left(\overline{A}\,B\right) = \left(A\,\overline{A}\right)B = \{\emptyset\}B = \{\emptyset\}$, then A and $\overline{A}\,B$ are mutually exclusive events, as required.

- Second, and using a similar approach, note that:

$$B = S\,B = \left(A \cup \overline{A}\right)B = (A\,B) \cup \left(\overline{A}\,B\right) \qquad \square$$

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

Set Theory

Example (Proof of the Sum Rule). SOLUTION. Using these two disjoint unions, then:

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

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

□

Set Theory

Example (Proof of the Sum Rule). SOLUTION. Using these two disjoint unions, then:

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

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

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

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

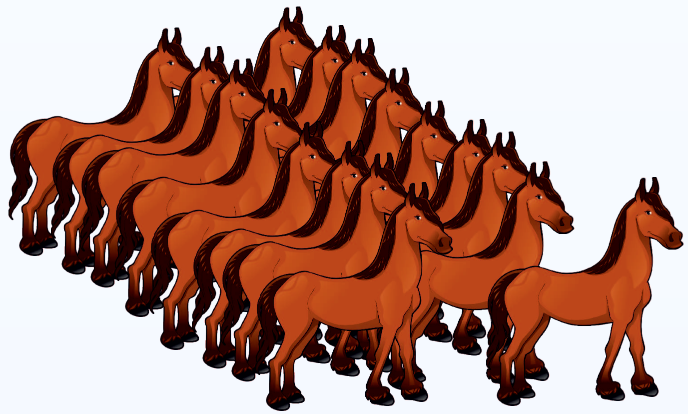
Set Theory

– End-of-Topic 14: **Set theory and its used in probability theory.** –

Any Questions?

Countable Spaces and Total Probability

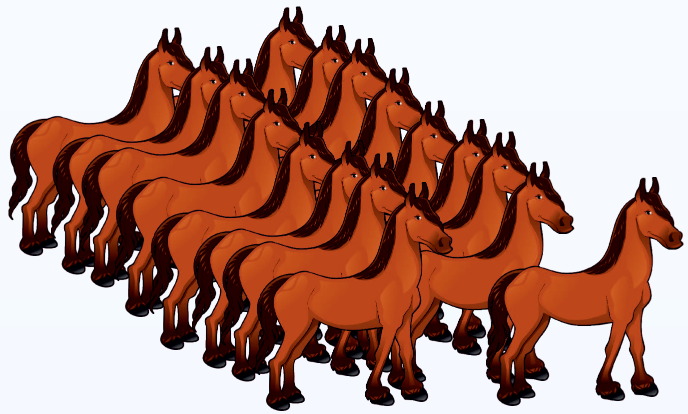
Example (Farmer and his Will). A farmer leaves a will saying that they wish for their first child to get half of his property, the second child to get a third, and the third child to get a ninth. As seventeen horses have been left, the children are distressed because they don't want to cut any horses up.



⋈

Countable Spaces and Total Probability

Example (Farmer and his Will). A farmer leaves a will saying that they wish for their first child to get half of his property, the second child to get a third, and the third child to get a ninth. As seventeen horses have been left, the children are distressed because they don't want to cut any horses up.

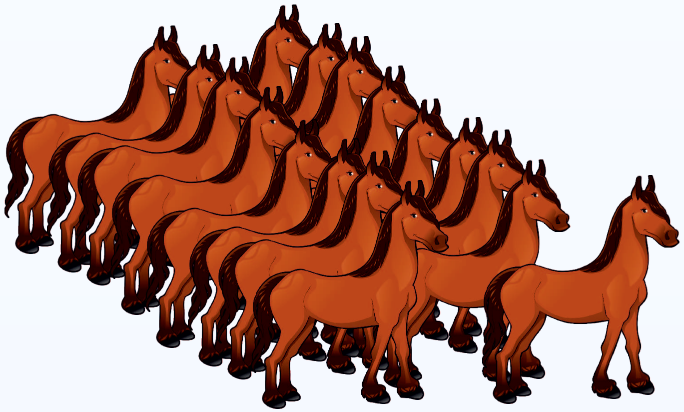


⋈

However, a local statistician lends them a horse so that they have eighteen. The children then take nine, six, and two horses, respectively. This adds up to seventeen, so they give the statistician the horse back, and everyone is happy.

Countable Spaces and Total Probability

Example (Farmer and his Will). A farmer leaves a will saying that they wish for their first child to get half of his property, the second child to get a third, and the third child to get a ninth. As seventeen horses have been left, the children are distressed because they don't want to cut any horses up.



⋈

However, a local statistician lends them a horse so that they have eighteen. The children then take nine, six, and two horses, respectively. This adds up to seventeen, so they give the statistician the horse back, and everyone is happy.

What is wrong with this story?

Countable Spaces and Total Probability

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 $\{\zeta_i\}$.

From the basic axioms, it follows that $p_i \geq 0$ and that

$$\sum_{i=1}^{N} p_i = 1$$

Countable Spaces and Total Probability

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 $\{\zeta_i\}$.

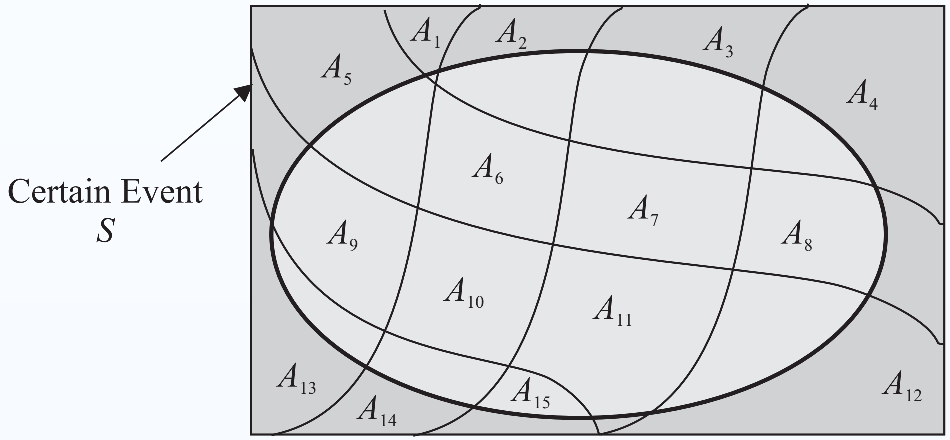
From the basic axioms, it follows that $p_i \geq 0$ and that

$$\sum_{i=1}^{N} p_i = 1$$

- This can be used in obtaining the **principle of total probability**.
- Let $A_1, A_2, A_3, \dots$ be a finite or countably infinite set of mutually exclusive and collectively exhaustive events, then

$$\sum_i \Pr(A_i \cap B) = \Pr(B)$$

Countable Spaces and Total Probability



- This can be used in obtaining the **principle of total probability**.
- Let $A_1, A_2, A_3, \dots$ be a finite or countably infinite set of mutually exclusive and collectively exhaustive events, then

$$\sum_i \Pr(A_i \cap B) = \Pr(B)$$

Countable Spaces and Total Probability

After this lecture, try the following example in the notes:

Example (Detection and Classification). An acoustic scene analysis algorithm is monitoring animal sounds, and makes sound classifications, either being labelled as bird, fox, or pet sounds.

- 29% of the detected sounds are false alarms;
- 3% of labelled bird sounds are false alarm detections;
- 12% of detected bird sounds are correctly labelled;
- 5% of labelled fox sounds are false alarm detections;
- 32% are correct detections of domestic pet sounds.

The following events are defined: correctly classified – C; mis-classified – M; bird sound – B; fox sound – F; pets – D.

Countable Spaces and Total Probability

After this lecture, try the following example in the notes:

Example (Detection and Classification). An acoustic scene analysis algorithm is monitoring animal sounds, and makes sound classifications, either being labelled as bird, fox, or pet sounds.

Draw a Venn diagram of the problem, and determine:

1. What is the probability that a detection is classified as a bird sound, either correctly or incorrectly?
2. What is the probability that a detection is a false alarm and/or a labelled bird sound?
3. What is the probability that a sound is correctly classified as a fox or domestic pet sound?
4. What is the probability of a false alarm for a pet sound?

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.

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.

Suppose that S is the set of all real numbers. 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.

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.

Suppose that S is the set of all real numbers. 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 $\{x \leq x_i\}$.

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.

Suppose that S is the set of all real numbers. 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.

The Real Line

– End-of-Topic 15: **Countable Spaces, Total Probabilities, and Uncountable Spaces on the Real line** –

Any Questions?

Conditional Probability

If an experiment is repeated n times, and on each occasion the occurrences or non-occurrences two events A and B are observed. Suppose that only those outcomes for which B occurs are considered, and all other experiments are disregarded.

Conditional Probability

If an experiment is repeated n times, and on each occasion the occurrences or non-occurrences 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:

$$\Pr\left(A \,\middle|\, B\right) \approx \frac{n_{AB}}{n_B} = \frac{n_{AB}/n}{n_B/n} = \frac{\Pr\left(AB\right)}{\Pr\left(B\right)}$$

provided that n is sufficiently large.

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

Conditional Probability

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

Conditional Probability

Example (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.

C_1	C_2	Outcome	
Gender	Gender	Relevant?	Desired?
B	B	✓	✓
G	B	✓	
B	G	✓	
G	G		
Count		3	1

□

Therefore, using classical probability, since the events are all equally probable, the answer is $p = {}^{N_A}/_{N} = {}^1/_3$.

Bayes's Rule

Conditional probability leads onto Bayes's theorem.

$$\Pr\left(AB\right) = \Pr\left(A \,\middle|\, B\right) \Pr\left(B\right) = \Pr\left(B \,\middle|\, A\right) \Pr\left(A\right)$$

Bayes's Rule

Conditional probability leads onto Bayes's theorem.

$$\Pr\left(AB\right) = \Pr\left(A \,\middle|\, B\right) \Pr\left(B\right) = \Pr\left(B \,\middle|\, A\right) \Pr\left(A\right)$$

giving

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

Bayes's Rule

Conditional probability leads onto Bayes's theorem.

$$\Pr(AB) = \Pr\left(A \,\middle|\, B\right) \Pr(B) = \Pr\left(B \,\middle|\, A\right) \Pr(A)$$

giving

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

- Bayes's rule will be used throughout this course, and commonly arises in the analysis of signal and communication systems, machine learning, and data science.
- Bayesian inference is typically a computationally expensive problem, but can be solved efficiently using graphical models, sparsity, and numerical Bayesian methods such as Monte Carlo and Message Passing techniques.

Bayes's Rule

Example (Prisoner's Problem). Three prisoners, A, B and C, are in separate cells. The governor has selected one of them at random to be pardoned. The warden knows which one is to be released, but is not allowed to say. Prisoner A begs the warden to be told the identity of one of the *others* who **will not** be released.

> Prisoner A says: *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 will not be released.

Bayes's Rule

Example (Prisoner's Problem). Three prisoners, A, B and C, are in separate cells. The governor has selected one of them at random to be pardoned. The warden knows which one is to be released, but is not allowed to say. Prisoner A begs the warden to be told the identity of one of the *others* who **will not** be released.

Prisoner A says: *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 will not be released.

Prisoner A believes that the probability of being released has gone up from $^1/_3$ to $^1/_2$, as it is now between A and C. Prisoner A tells C the news, who reasons that A still has a chance of $^1/_3$ to be the pardoned one, but C's chance has gone up to $^2/_3$. What is the correct answer?

Bayes's Rule

Example (Prisoner's Problem). Prisoner A says: *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 will not be released.

SOLUTION. Solve using total probability and Bayes's theorem.

- Let A, B, and C be the events that the corresponding prisoner will be pardoned.
- Note that A, B, and C are independent events, *before* the warden has provided any information.
- Let b be the event that the warden tells A that B is **not** to be released.

Bayes's Rule

Example (Prisoner's Problem). Prisoner A says: *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 will not be released.

SOLUTION. Using Bayes's theorem, it follows that:

$$\Pr\left(A \,\middle|\, b\right) = \frac{\Pr\left(b \,\middle|\, A\right)\,\Pr\left(A\right)}{\Pr\left(b\right)}$$

□

Bayes's Rule

Example (Prisoner's Problem). Prisoner A says: *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 will not be released.

SOLUTION. Using the principal of total probability:

$$\begin{aligned}
\Pr(b) &= \sum_{i\in\{A,B,C\}} \Pr(b,i) \\
&= \Pr(b,A) + \Pr(b,B) + \Pr(b,C) \\
&= \Pr(b \mid A)\Pr(A) + \Pr(b \mid B)\Pr(B) + \Pr(b \mid C)\Pr(C) \\
&= \frac{1}{2}\times\frac{1}{3} + 0\times\frac{1}{3} + 1\times\frac{1}{3} = \frac{1}{2}
\end{aligned}$$

□

Bayes's Rule

Example (Prisoner's Problem). Prisoner A says: *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 will not be released.

SOLUTION.
- If A is to be released, the warden can tell A either B or C through the toss of the coin $\Rightarrow \Pr\left(b \,\middle|\, A\right) = \frac{1}{2}$.
- If C is to be released, the warden is now constraned to say B will not be released, so $\Pr\left(b \,\middle|\, C\right) = 1$. □

Bayes's Rule

Example (Prisoner's Problem). Prisoner A says: *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 will not be released.

SOLUTION.
- If A is to be released, the warden can tell A either B or C through the toss of the coin $\Rightarrow \Pr\left(b \,|\, A\right) = \frac{1}{2}$.
- If C is to be released, the warden is now constraned to say B will not be released, so $\Pr\left(b \,|\, C\right) = 1$.

$$\Pr\left(A \,|\, b\right) = \frac{\Pr\left(b \,|\, A\right)\, \Pr\left(A\right)}{\Pr\left(b\right)} = \frac{\frac{1}{2} \times \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}$$

□

Bayes's Rule

Example (Prisoner's Problem). Prisoner A says: *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 will not be released.

SOLUTION. ● If A is to be released, the warden can tell A either B or C through the toss of the coin $\Rightarrow \Pr\left(b \mid A\right) = \frac{1}{2}$.

● If C is to be released, the warden is now constraned to say B will not be released, so $\Pr\left(b \mid C\right) = 1$.

$$\Pr\left(A \mid b\right) = \frac{\Pr\left(b \mid A\right)\, \Pr\left(A\right)}{\Pr\left(b\right)} = \frac{\frac{1}{2} \times \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}$$

$$\Pr\left(C \mid b\right) = \frac{\Pr\left(b \mid C\right)\, \Pr\left(C\right)}{\Pr\left(b\right)} = \frac{1 \times \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

Bayes's Rule

Example (Prisoner's Problem). Prisoner A says: *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 will not be released.

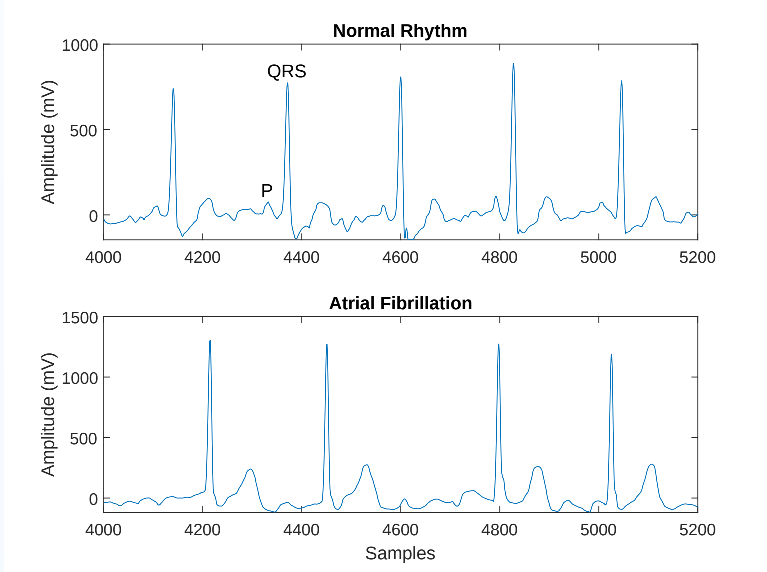
SOLUTION. The tendency of people to provide the answer $1/2$ neglects to take into account that the warden may have tossed a coin before giving an answer. The warden may have answered B because either:

- A is to be released and the wardan tossed a coin;
- or C is to be released.

The probabilities of these two events are not equal.

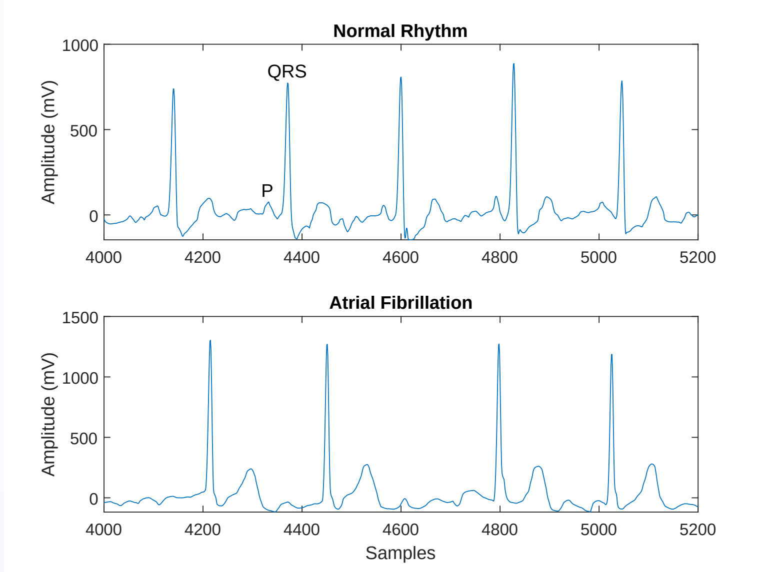
Bayes's Rule

After this lecture, try the following example in the notes:



Example (Classification Accuracy). An algorithm using electrocardiogram (ECG) data is used to test for a certain irregular heartbeat and is 95% accurate. A person submits to the test and the results are positve. Suppose the person comes from a population of 10^5, where 2000 people suffer the irregularity.

Bayes's Rule



Example (Classification Accuracy). An algorithm using electrocardiogram (ECG) data is used to test for a certain irregular heartbeat and is 95% accurate. A person submits to the test and the results are positve. Suppose the person comes from a population of 10^5, where 2000 people suffer the irregularity.

What can we conclude about the probability that the person under test has that particular heartbeat irregularity?

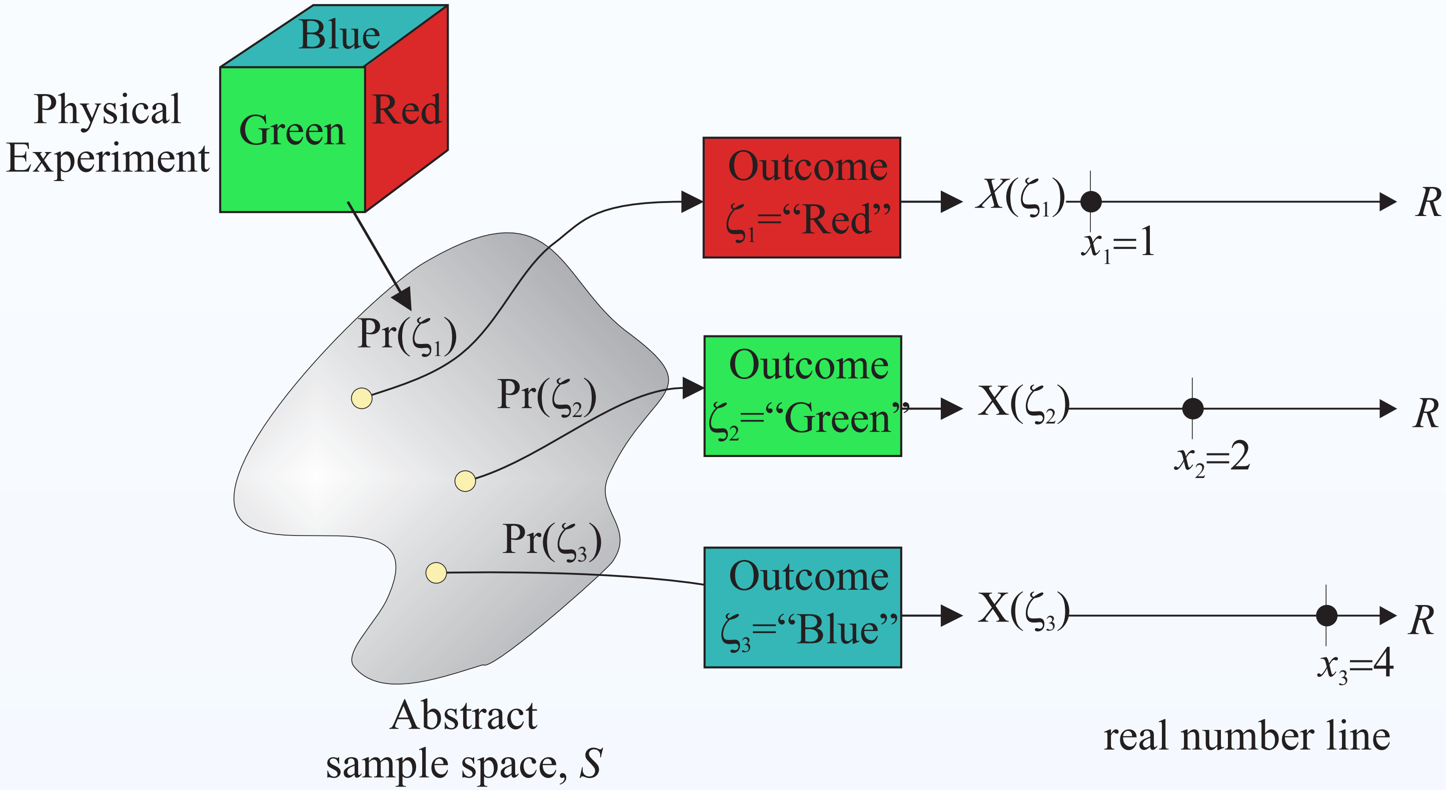
Bayes's Rule

– End-of-Topic 16: **Conditional Probability, and a basic but important Introduction to Bayes Rule** –

Any Questions?

Lecture Slideset 2
Scalar Random Variables

Definition



A graphical representation of a random variable for a more specific example.

- **Note that for continuous random variables, the outcomes are events, such as small intervals on the real axis as described in the previous lecture.**

Definition

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.

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 $\Pr(X(\zeta) = -\infty) = 0$.

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

Definition

Example (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:

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

$\bowtie$

Definition

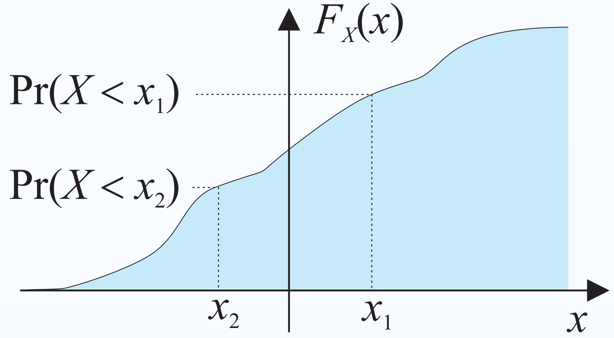
Example (Rolling die). Consider rolling a die, with six outcomes $\{\zeta_i,\, i \in \{1, \dots, 6\}\}$. In this experiment, assign the number 1 to every *even* outcome, and the number 0 to every *odd* outcome. Then the **RV** $X(\zeta)$ is given by:

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

⋈

Example (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 $X(\zeta) \le 5$ corresponds to the letters A, B, C, D, or E.

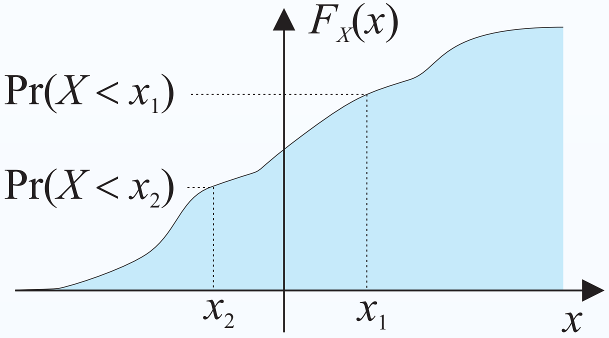
Distribution functions



The cumulative distribution function.

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

Distribution functions

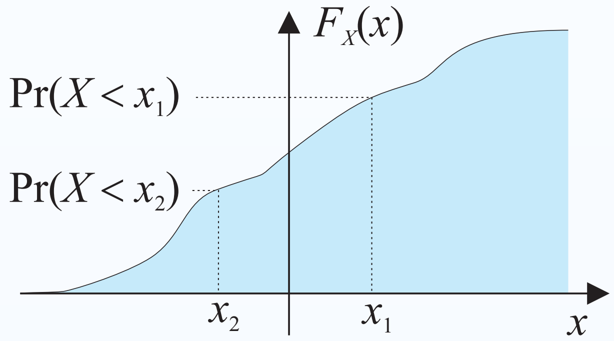


The cumulative distribution function.

- 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:

$$F_X(x) \triangleq \Pr(X(\zeta) \leq x)$$

Distribution functions

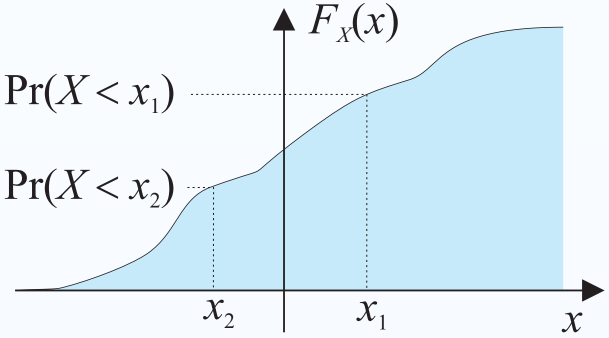


The cumulative distribution function.

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

$$\begin{aligned}\Pr\left(x_\ell < X(\zeta) \le x_r\right) &= \Pr\left(X(\zeta) \le x_r\right) - \Pr\left(X(\zeta) \le x_\ell\right) \\ &= F_X\left(x_r\right) - F_X\left(x_\ell\right)\end{aligned}$$

Distribution functions



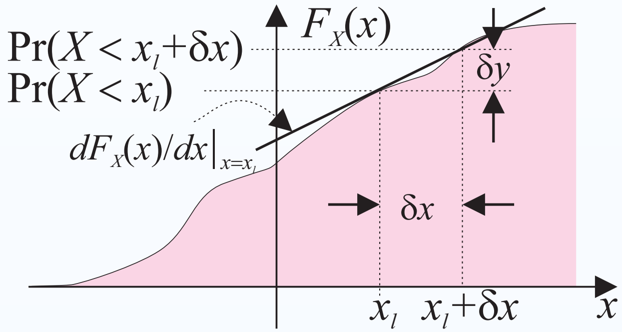
The cumulative distribution function.

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

$$\begin{aligned}\Pr\left(x_\ell < X(\zeta) \le x_r\right) &= \Pr\left(X(\zeta) \le x_r\right) - \Pr\left(X(\zeta) \le x_\ell\right)\\ &= F_X\left(x_r\right) - F_X\left(x_\ell\right)\end{aligned}$$

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

Distribution functions



The gradient of the cdf is important, and leads to the pdf.

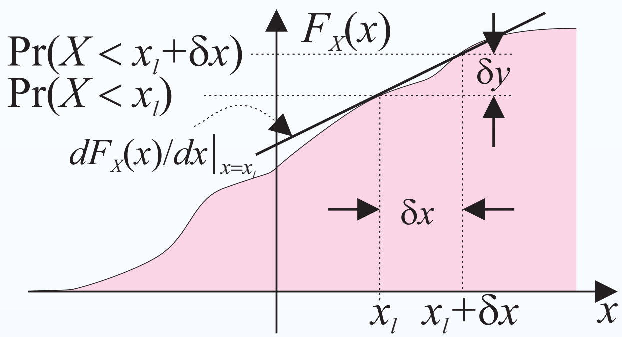
This can be seen by setting $x_r = x_l + \delta x$:

$$\Pr\left(x_\ell < X(\zeta) \le x_\ell + \delta x\right) = \Pr\left(X(\zeta) \le x_\ell + \delta x\right) - \Pr\left(X(\zeta) \le x_\ell\right)$$

$$\approx \Pr\left(X(\zeta) \le x_\ell\right) + \left.\frac{dF_X(x)}{dx}\right|_{x=x_\ell} \delta x - \Pr\left(X(\zeta) \le x_\ell\right)$$

$$\approx \left.\frac{dF_X(x)}{dx}\right|_{x=x_\ell} \delta x$$

Distribution functions



The gradient of the cdf is important, and leads to the pdf.

This can be seen by setting $x_r = x_l + \delta x$:

$$\Pr(x_\ell < X(\zeta) \le x_\ell + \delta x) = \Pr(X(\zeta) \le x_\ell + \delta x) - \Pr(X(\zeta) \le x_\ell)$$

$$\approx \Pr(X(\zeta) \le x_\ell) + \left.\frac{dF_X(x)}{dx}\right|_{x=x_\ell} \delta x - \Pr(X(\zeta) \le x_\ell)$$

$$\approx \left.\frac{dF_X(x)}{dx}\right|_{x=x_\ell} \delta x$$

Shortly, it will be seen that $\frac{dF_X(x)}{dx}$ is indeed the pdf.

Kolmogorov's Axioms

The events $\{X(\zeta) \le x_1\}$ and $\{x_1 < X(\zeta) \le x_2\}$ are mutually exclusive events. Therefore, their union equals $\{X(\zeta) \le x_2\}$, and thus:

$$\Pr\left(X(\zeta) \le x_1\right) + \Pr\left(x_1 < X(\zeta) \le x_2\right) = \Pr\left(X(\zeta) \le x_2\right)$$

$$\int_{-\infty}^{x_1} p(v)\, dv + \Pr\left(x_1 < X(\zeta) \le x_2\right) = \int_{-\infty}^{x_2} p(v)\, dv$$

$$\Rightarrow \qquad \Pr\left(x_1 < X(\zeta) \le x_2\right) = \int_{x_1}^{x_2} p(v)\, dv$$

where $p(v)$ is an probability density function (pdf) that will be described in more detail in the next section.

Moreover, it follows that $\Pr\left(-\infty < X(\zeta) \le \infty\right) = 1$ and the probability of the impossible event, $\Pr\left(X(\zeta) \le -\infty\right) = 0$. Hence, the cdf satisfies the axiomatic definition of probability.

Kolmogorov's Axioms

– End-of-Topic 17: **Introduction to Random Variables and Cummulative Distribution Functions** –

Any Questions?

Density functions

It was seen in the previous section that gradients of the cdf are important when determining the probability of being within small intervals.

- The **probability density function (pdf)** of a **RV**, $X(\zeta)$, is:

$$f_X(x) \triangleq \frac{dF_X(x)}{dx}$$

Note $f_X(x)$ is not a **probability** on its own; it must be multiplied by a certain interval Δx to obtain a probability:

$$f_X(x)\,\Delta x \approx F_X(x+\Delta x) - F_X(x) \approx \Pr(x < X(\zeta) \le x + \Delta x)$$

Density functions

It was seen in the previous section that gradients of the cdf are important when determining the probability of being within small intervals.

- The **probability density function (pdf)** of a **RV**, $X(\zeta)$, is:

$$f_X(x) \triangleq \frac{dF_X(x)}{dx}$$

Note $f_X(x)$ is not a **probability** on its own; it must be multiplied by a certain interval Δx to obtain a probability:

$$f_X(x)\,\Delta x \approx F_X(x+\Delta x) - F_X(x) \approx \Pr\left(x < X(\zeta) \leq x + \Delta x\right)$$

- It directly follows that:

$$F_X(x) = \int_{-\infty}^{x} f_X(v)\, dv$$

Density functions

- For discrete-valued **RV**, use the **probability mass function (pmf)**, p_k, the probability that $X(\zeta)$ takes on a value equal to x_k: $p_k \triangleq \Pr\left(X(\zeta) = x_k\right)$.

Density functions

- For discrete-valued **RV**, use the **probability mass function (pmf)**, p_k, the probability that $X(\zeta)$ takes on a value equal to x_k: $p_k \triangleq \Pr\left(X(\zeta) = x_k\right)$.

The pmf for a discrete RVs can be written as a pdf through:

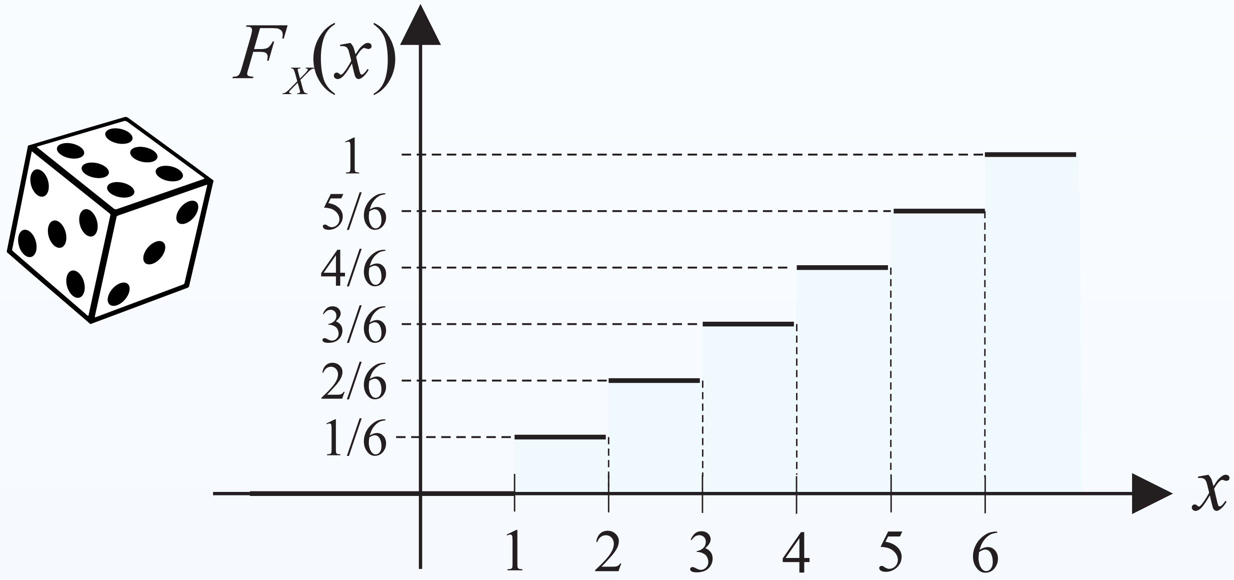
$$f_X\left(x\right) = \sum_k p_k \, \delta(x - x_k)$$

where $\delta(x)$ is the Dirac-delta function, and is given by:

$$\delta(x) = 0 \qquad \text{if } x \neq 0$$

$$\int_{-\infty}^{\infty} \delta(x)\ dx = 1$$

Density functions



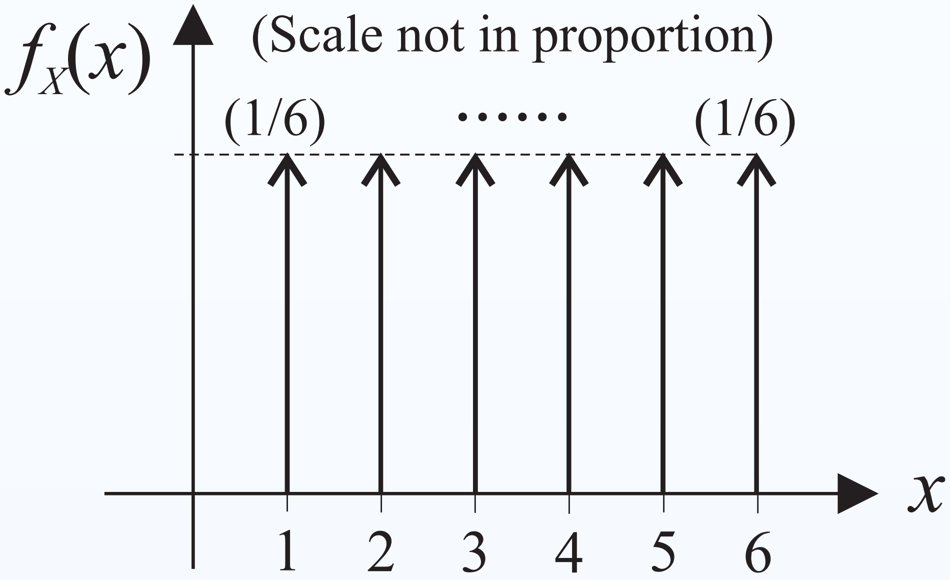
The cdf and pdf for a fair six-sided die.

Example (die). Describe the cdf and pdf for a fair six-sided die.

SOLUTION. The probability mass function (pmf) is given by $p_i = \Pr\left(X(\zeta) = x_i\right) = \frac{1}{6}$, where $x_i = i$, $i \in \{1, \dots, 6\}$.

Note that $\Pr\left(X(\zeta) < x_1\right) = 0$ whereas $\Pr\left(X(\zeta) \leq x_1\right) = {}^1\!/_6$.

Density functions



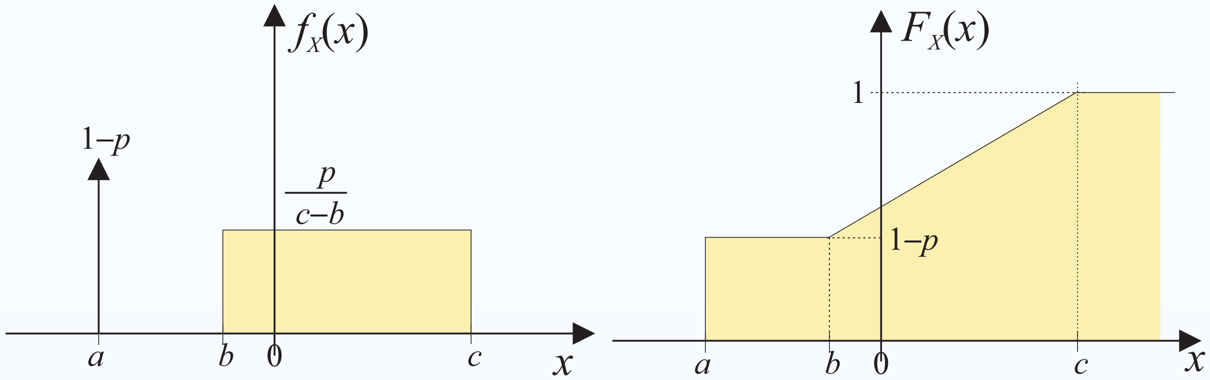
The cdf and pdf for a fair six-sided die.

Example (die). Describe the cdf and pdf for a fair six-sided die.

SOLUTION. The pdf is obtained by differentiating the cdf:

$$f_X(x) = \sum_{i=1}^{N} p_i\,\delta(x - x_i) = \frac{1}{6}\sum_{i=1}^{6}\delta(x - i) \qquad \square$$

Density functions

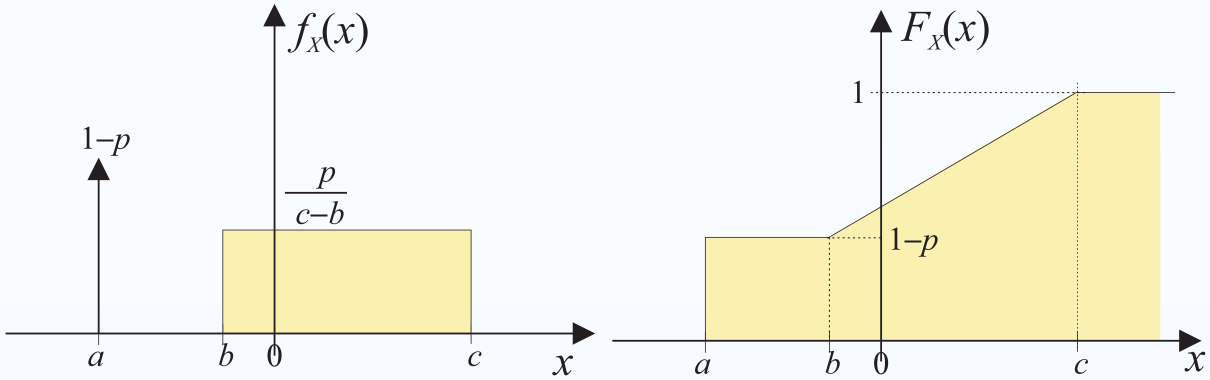


A probability density function and its corresponding cumulative distribution function for a RV which is a mixture of continuous and discrete components.

Moreover, a mixture of continuous and discrete components will have a pdf composed of delta as well as continous functions:

$$f_{X,m}(x) = \sum_k p_k\, \delta(x - x_k) + f_{X,c}(x)$$

Density functions



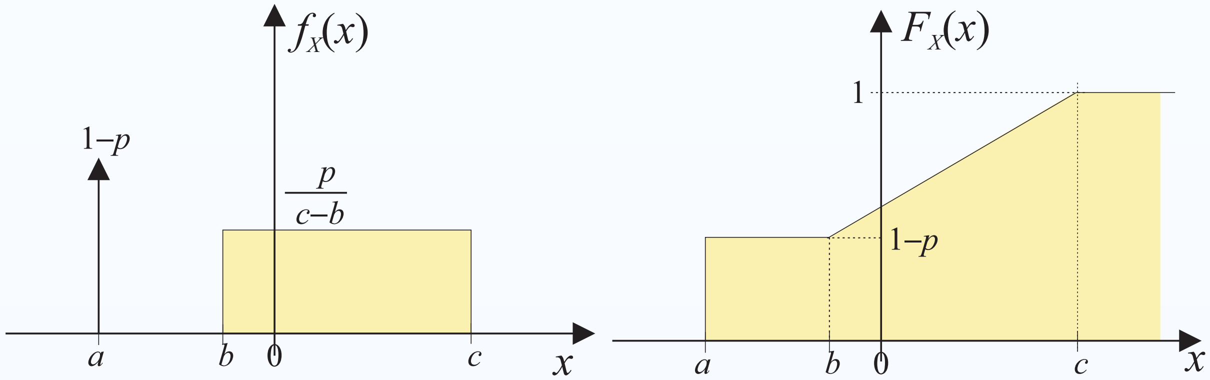
A probability density function and its corresponding cumulative distribution function for a RV which is a mixture of continuous and discrete components.

The pdf for the distribution shown above can be written as:

$$f_X(x) = (1-p)\,\delta(x-a) + \frac{p}{c-b}\left(u(x-b) - u(x-c)\right)$$

where $u(x)$ is the unit step function, such that $u(x) = 1$ if $x \geq 0$ and zero otherwise.

Density functions

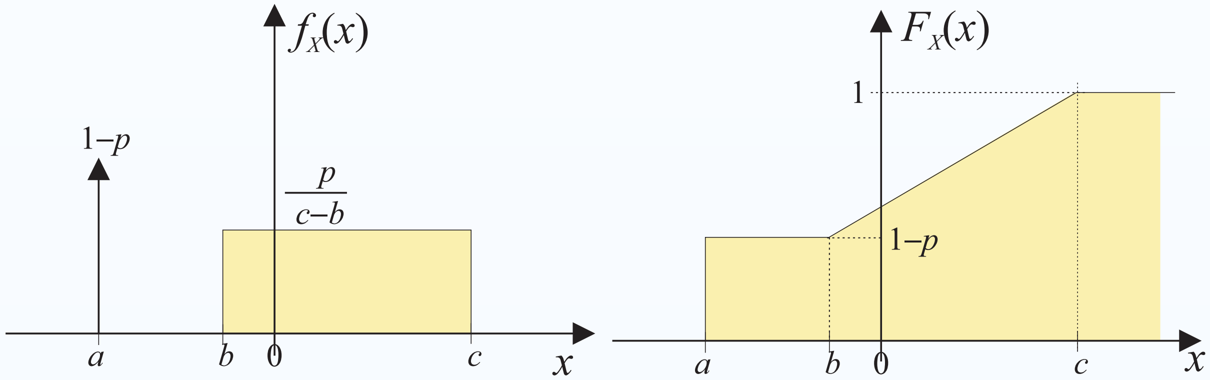


A probability density function and its corresponding cumulative distribution function for a RV which is a mixture of continuous and discrete components.

Integrating, it is can be shown that:

$$F_X(\infty) = \int_{-\infty}^{\infty} f_X(x)\, dx = (1-p) + \frac{p}{c-b} \times (c-b) = 1$$

Density functions



A probability density function and its corresponding cumulative distribution function for a RV which is a mixture of continuous and discrete components.

Integrating, it is can be shown that:

$$F_X(\infty) = \int_{-\infty}^{\infty} f_X(x)\, dx = (1-p) + \frac{p}{c-b} \times (c-b) = 1$$

Can you think of examples of a mixture of discrete and continuous random variables?

Properties: Distributions and Densities

- Properties of **cdf**:

$$0 \leq F_X(x) \leq 1, \quad \lim_{x \to -\infty} F_X(x) = 0, \quad \lim_{x \to \infty} F_X(x) = 1$$

$F_X(x)$ is a monotonically increasing function of x:

$$F_X(a) \leq F_X(b) \quad \text{if} \quad a \leq b$$

Properties: Distributions and Densities

- Properties of **cdf**:

$$0 \leq F_X(x) \leq 1, \quad \lim_{x \to -\infty} F_X(x) = 0, \quad \lim_{x \to \infty} F_X(x) = 1$$

$F_X(x)$ is a monotonically increasing function of x:

$$F_X(a) \leq F_X(b) \quad \text{if} \quad a \leq b$$

- Properties of **pdfs**:

$$f_X(x) \geq 0, \quad \int_{-\infty}^{\infty} f_X(x)\, dx = 1$$

Properties: Distributions and Densities

- Properties of **cdf**:

$$0 \leq F_X(x) \leq 1, \quad \lim_{x \to -\infty} F_X(x) = 0, \quad \lim_{x \to \infty} F_X(x) = 1$$

$F_X(x)$ is a monotonically increasing function of x:

$$F_X(a) \leq F_X(b) \quad \text{if} \quad a \leq b$$

- Properties of **pdfs**:

$$f_X(x) \geq 0, \quad \int_{-\infty}^{\infty} f_X(x)\, dx = 1$$

- Probability of arbitrary events:

$$\Pr(x_1 < X(\zeta) \leq x_2) = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(x)\, dx$$

Properties: Distributions and Densities

– End-of-Topic 18: **Introduction to pdf and their properties** –

Any Questions?

Common Continuous RVs

Uniform distribution

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x \leq b, \\ 0 & \text{otherwise} \end{cases}$$

Normal distribution

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right], \quad x \in \mathbb{R}$$

Cauchy distribution

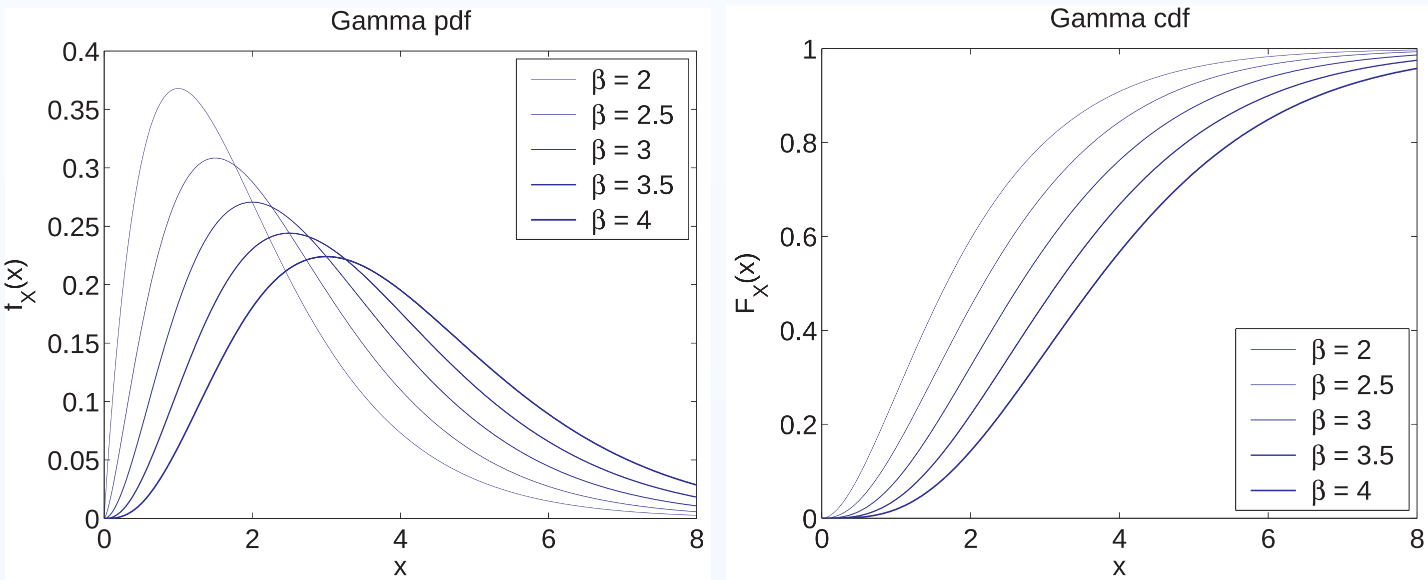
$$f_X(x) = \frac{\beta}{\pi} \frac{1}{(x-\mu_X)^2 + \beta^2}$$

The Cauchy random variable is symmetric around the value $x = \mu_X$, but its mean and variance do not exist.

Common Continuous RVs

Gamma distribution

$$f_X(x) = \begin{cases} 0 & \text{if } x < 0, \\ \frac{1}{\Gamma(\beta)} \alpha^\beta \, x^{\beta-1} \, e^{-\alpha x} & \text{if } x \geq 0, \end{cases}$$

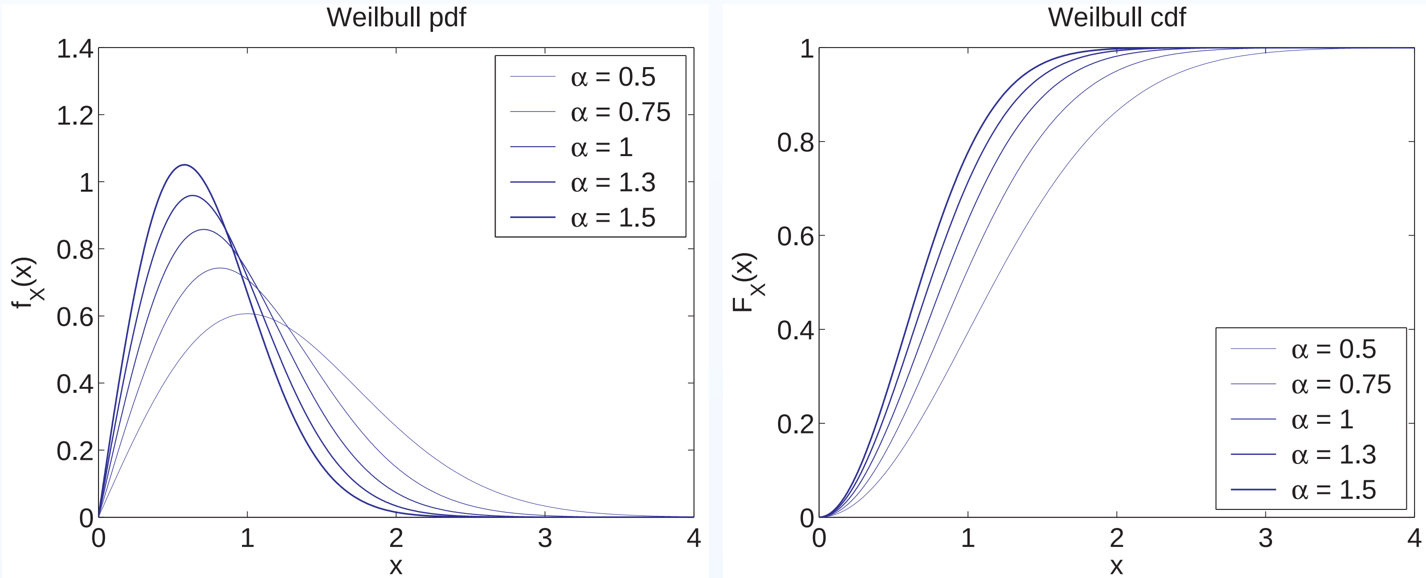


The Gamma density and distribution functions, for the case when $\alpha = 1$ and for various values of β.

Common Continuous RVs

Weibull distribution

$$f_X(x) = \begin{cases} 0 & x < 0 \\ \alpha\beta x^{\beta-1}\, e^{-\alpha\, x^{\beta}} & x \geq 0 \end{cases}$$



The Weibull density and distribution functions, for the case when $\alpha = 1$, and for various values of the parameter β.

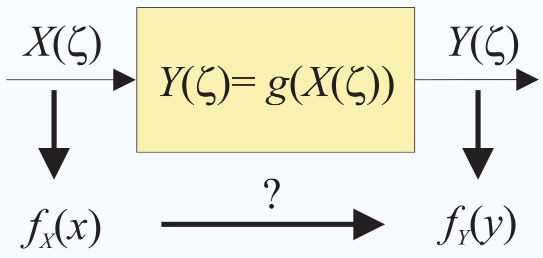
Common Continuous RVs

– End-of-Topic 19: **Introduction to common density functions** –

Any Questions?

Probability transformation rule

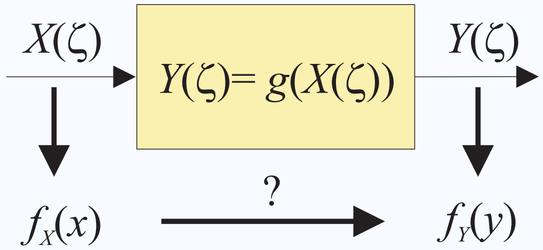
Suppose a random variable $Y(\zeta)$ is a function, g, of a random variable $X(\zeta)$, which has pdf given by $f_X(x)$. What is $f_Y(y)$?



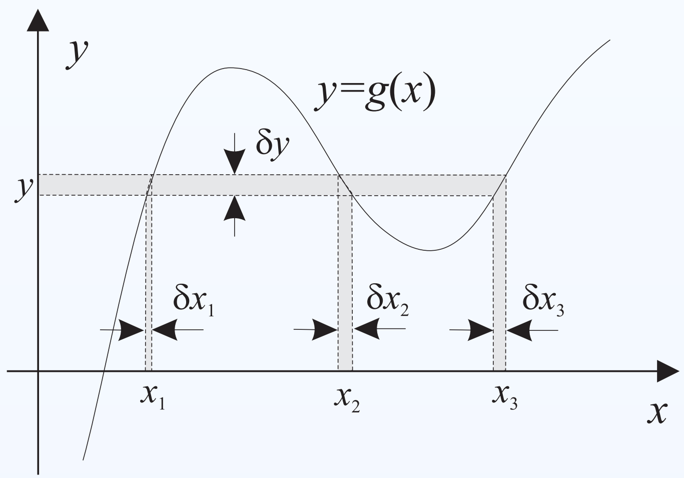
The mapping $y = g(x)$.

Probability transformation rule

Suppose a random variable $Y(\zeta)$ is a function, g, of a random variable $X(\zeta)$, which has pdf given by $f_X(x)$. What is $f_Y(y)$?

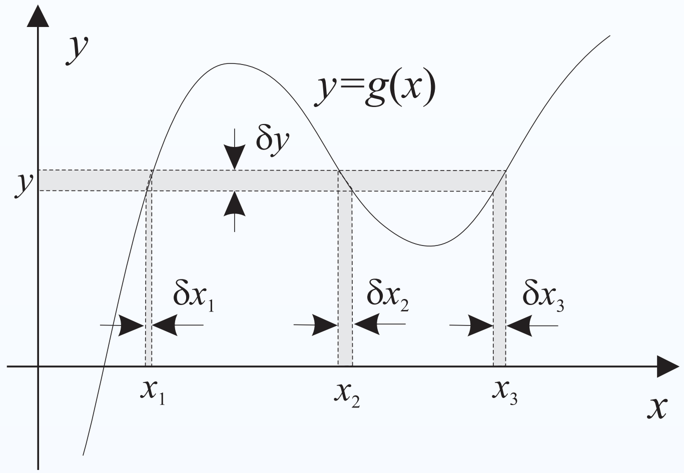


The mapping $y = g(x)$**.**



The mapping $y = g(x)$**.**

Probability transformation rule

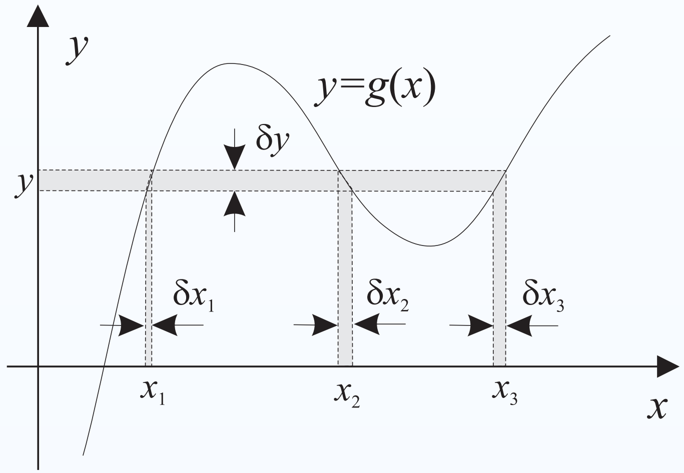


Theorem (Probability Transformation). PROOF. First consider the output **pdf** which, by definition, is given by:

$$f_Y(y)\, dy = \Pr\left(y < Y(\zeta) \leq y + dy\right)$$

□

Probability transformation rule



Theorem (Probability Transformation). PROOF. First consider the output **pdf** which, by definition, is given by:

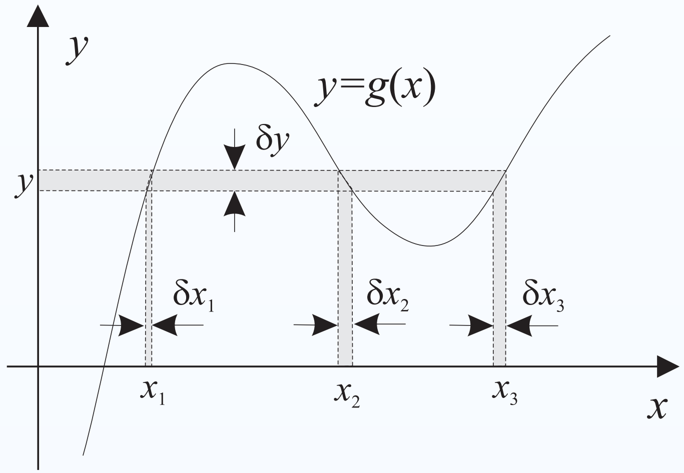
$$f_Y(y)\; dy = \Pr\left(y < Y(\zeta) \leq y + dy\right)$$

The set of values x such that $y < g(x) \leq y + dy$ consists of the intervals:

$$x_n < x \leq x_n + dx_n$$

□

Probability transformation rule

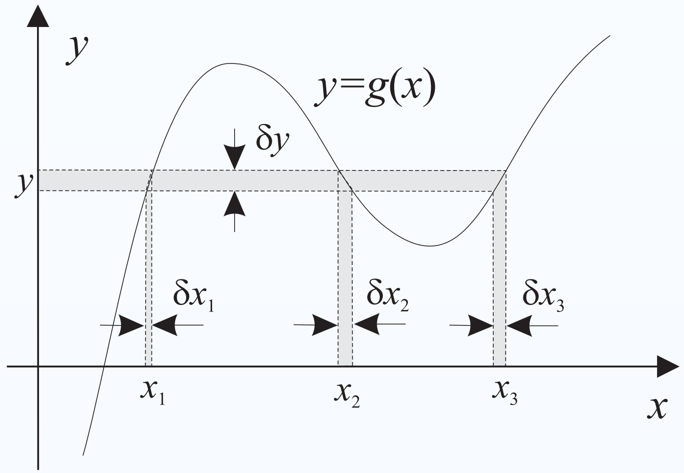


Theorem (Probability Transformation). PROOF. The probability that x lies in this set is

$$f_X(x_n)\, dx_n = \Pr\left(x_n < X(\zeta) \leq x_n + dx_n\right)$$

□

Probability transformation rule



Theorem (Probability Transformation). PROOF. The probability that x lies in this set is

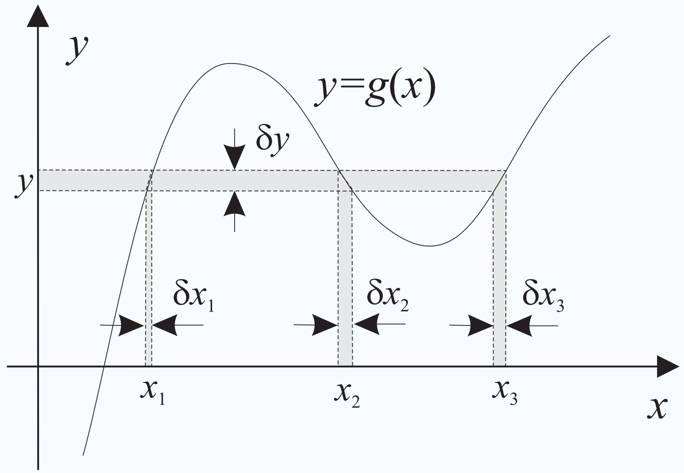
$$f_X(x_n)\,dx_n = \Pr\left(x_n < X(\zeta) \leq x_n + dx_n\right)$$

From the transformation from x to y, then

$$dx_n = \frac{dy}{|g'(x_n)|} \qquad \square$$

where $g'(x)$ is the derivative with respect to (w. r. t.) x of $g(x)$.

Probability transformation rule

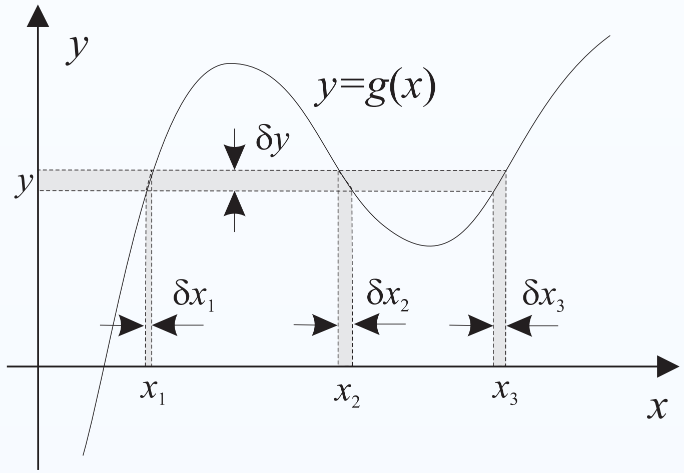


Theorem (Probability Transformation). PROOF. Finally, since these are N mutually exclusive sets, then

$$\Pr\left(y < Y(\zeta) \leq y + dy\right) = \sum_{n=1}^{N} \Pr\left(x_n < X(\zeta) \leq x_n + dx_n\right)$$

□

Probability transformation rule



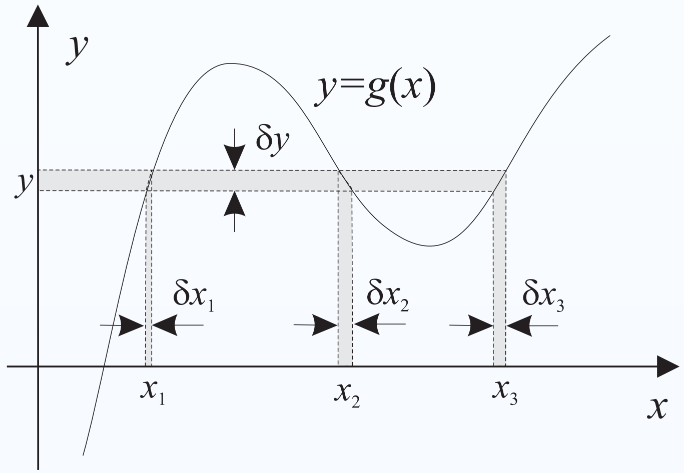
Theorem (Probability Transformation). PROOF. Finally, since these are N mutually exclusive sets, then

$$\Pr\left(y < Y(\zeta) \leq y + dy\right) = \sum_{n=1}^{N} \Pr\left(x_n < X(\zeta) \leq x_n + dx_n\right)$$

$$\approx\ f_Y\left(y\right)\ dy \approx \sum_{n=1}^{N} f_X\left(x_n\right)\ dx_n$$

□

Probability transformation rule



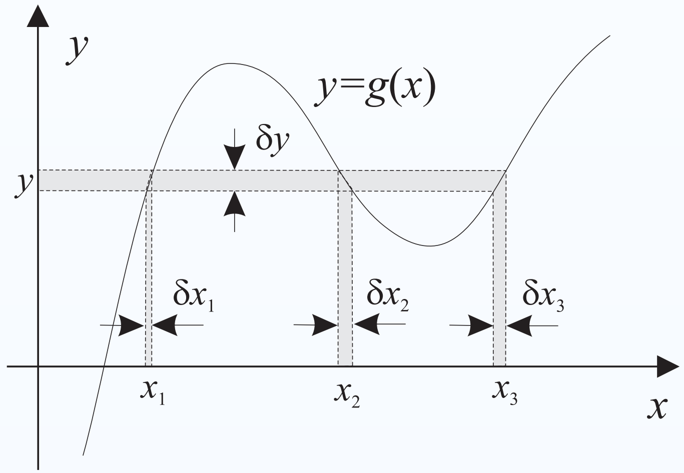
Theorem (Probability Transformation). PROOF. Finally, since these are N mutually exclusive sets, then

$$\Pr\left(y < Y(\zeta) \leq y + dy\right) = \sum_{n=1}^{N} \Pr\left(x_n < X(\zeta) \leq x_n + dx_n\right)$$

$$f_Y(y)\, dy = \sum_{n=1}^{N} f_X(x_n)\, \frac{dy}{|g'(x_n)|}$$

□

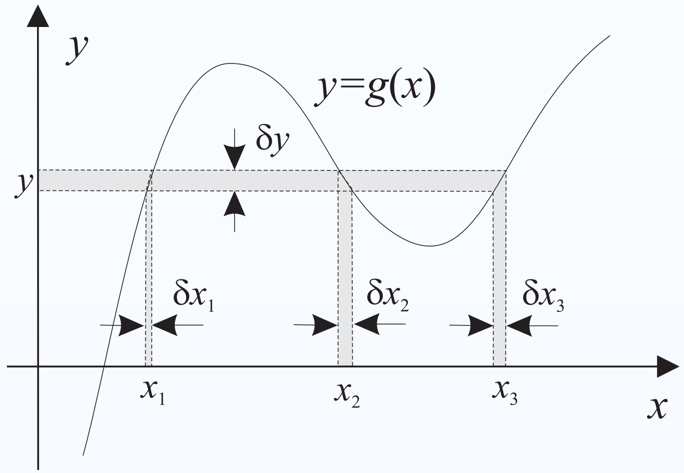
Probability transformation rule



Theorem (Probability Transformation). PROOF. Finally, since these are N mutually exclusive sets, then

$$f_Y(y) = \sum_{n=1}^{N} \frac{f_X(x_n)}{\left|\frac{dy}{dx}\right|_{x=x_n}}\Bigg|_{x_n=g^{-1}(y)} \qquad \square$$

Probability transformation rule

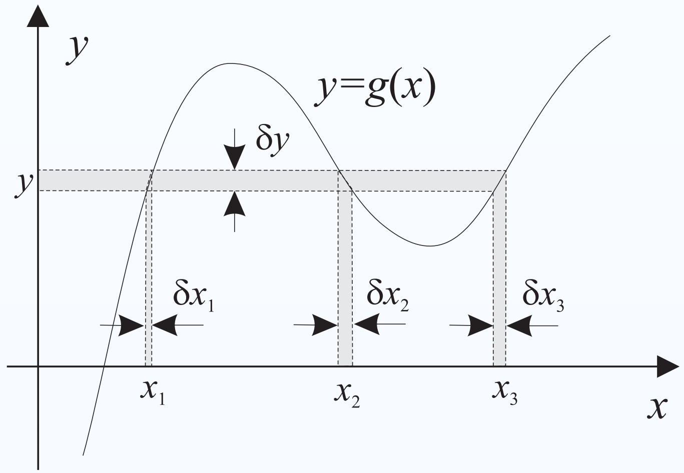


Theorem (Probability Transformation). Denote the real roots of $y = g(x)$ by $\{x_n,\ n \in \mathcal{N}\}$, such that:

$$y = g(x_1) = \cdots = g(x_N)$$

$\diamond$

Probability transformation rule



Theorem (Probability Transformation). Denote the real roots of $y = g(x)$ by $\{x_n,\, n \in \mathcal{N}\}$, such that:

$$y = g(x_1) = \cdots = g(x_N)$$

Then, if the $Y(\zeta) = g(X(\zeta))$, the pdf of $Y(\zeta)$ is given by:

$$f_Y(y) = \sum_{n=1}^{N} \frac{f_X(x_n)}{|g'(x_n)|}$$

Probability transformation rule

Example (Log-normal distribution). Let $Y = e^X$, where $X \sim \mathcal{N}(0, 1)$. Find the pdf for the RV Y.

Probability transformation rule

Example (Log-normal distribution). Let $Y = e^X$, where $X \sim \mathcal{N}(0, 1)$. Find the pdf for the RV Y.

SOLUTION. Since $X \sim \mathcal{N}(0, 1)$, then:

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

□

Probability transformation rule

Example (Log-normal distribution). Let $Y = e^X$, where $X \sim \mathcal{N}(0, 1)$. Find the pdf for the RV Y.

SOLUTION. Since $X \sim \mathcal{N}(0, 1)$, then:

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

- Considering the transformation $y = g(x) = e^x$, there is one root, given by $x = \ln y$.
- Therefore, the derivative of this expression is $g'(x) = \frac{d\, e^x}{dx} = e^x = y$. □

Probability transformation rule

Example (Log-normal distribution). Let $Y = e^X$, where $X \sim \mathcal{N}(0, 1)$. Find the pdf for the RV Y.

SOLUTION. Since $X \sim \mathcal{N}(0, 1)$, then:

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

- Considering the transformation $y = g(x) = e^x$, there is one root, given by $x = \ln y$.
- Therefore, the derivative of this expression is $g'(x) = \frac{d\, e^x}{dx} = e^x = y$.
- Hence, it follows:

$$f_Y(y) = \frac{f_X(x)}{g'(x)} = \frac{f_X(\ln y)}{y} = \frac{1}{y\sqrt{2\pi}} e^{-\frac{(\ln y)^2}{2}} \qquad \square$$

Probability transformation rule

After this lecture, try the following example in the notes:

Example (Inverse of a random variable). Let $Y = \frac{1}{X}$. Find the pdf for the RV Y, given by $f_Y(y)$, in terms of the pdf for the RV X, given by $f_X(x)$. Further, consider the special case when X has a **Cauchy density** with parameter α, such that:

$$f_X(x) = \frac{\alpha}{\pi} \frac{1}{x^2 + \alpha^2} \qquad \bowtie$$

Probability transformation rule

– End-of-Topic 20: **Derivation of the Probability Transformation Rule, and some examples** –

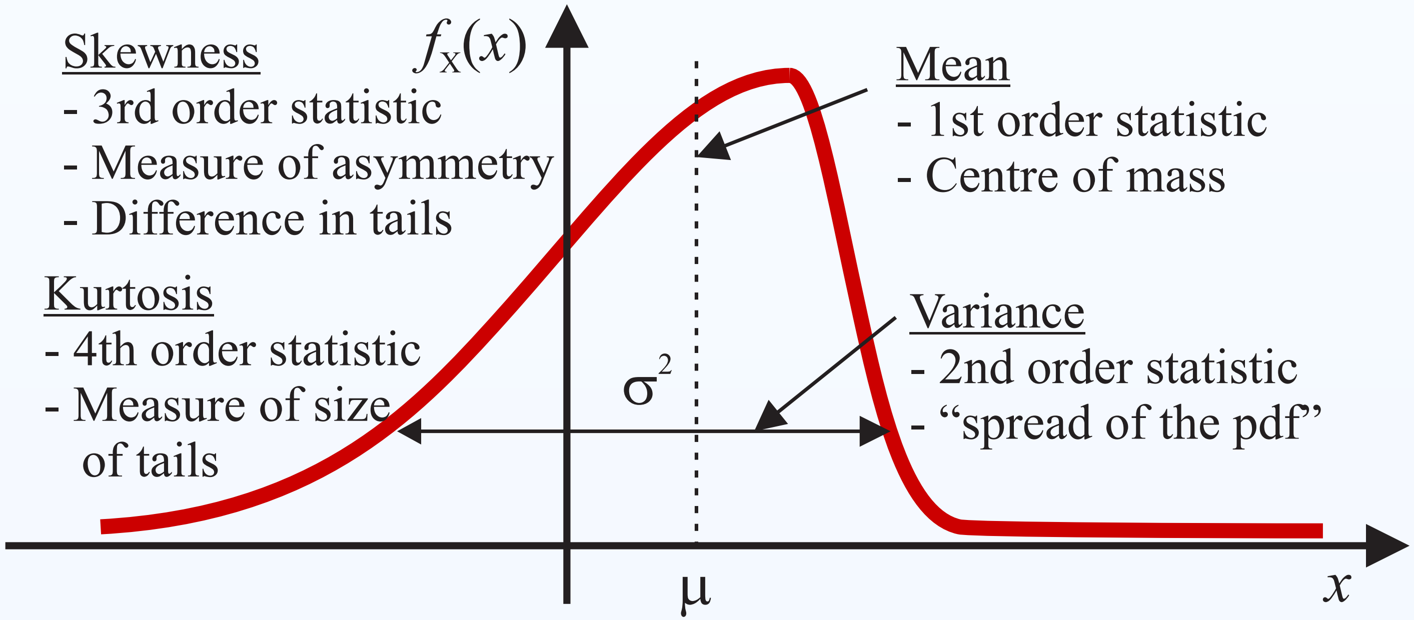
Any Questions?

Expectations

To completely characterise a **RV**, the **pdf** must be known. However, it is desirable to summarise key aspects of the **pdf** by using a few parameters rather than having to specify the entire density function.

Expectations

To completely characterise a **RV**, the **pdf** must be known. However, it is desirable to summarise key aspects of the **pdf** by using a few parameters rather than having to specify the entire density function.



The four salient or *key* features or *statistics of the pdf*.

Expectations

- The **expected** or **mean value** of a function of a **RV** $X(\zeta)$ is:

$$\mathbb{E}\left[X(\zeta)\right] = \int_{\mathbb{R}} x\, f_X(x)\; dx$$

Expectations

- The **expected** or **mean value** of a function of a **RV** $X(\zeta)$ is:

$$\mathbb{E}\left[X(\zeta)\right] = \int_{\mathbb{R}} x\, f_X\left(x\right)\, dx$$

- Recall: if $X(\zeta)$ is discrete then its corresponding **pdf** may be written in terms of its **pmf** as:

$$f_X\left(x\right) = \sum_k p_k\, \delta(x - x_k)$$

where the **Dirac-delta**, $\delta(x - x_k)$, is unity if $x = x_k$, and zero otherwise.

Expectations

- The **expected** or **mean value** of a function of a **RV** $X(\zeta)$ is:

$$\mathbb{E}\left[X(\zeta)\right] = \int_{\mathbb{R}} x\, f_X\left(x\right)\, dx$$

- Recall: if $X(\zeta)$ is discrete then its corresponding **pdf** may be written in terms of its **pmf** as:

$$f_X\left(x\right) = \sum_k p_k\, \delta(x - x_k)$$

where the **Dirac-delta**, $\delta(x - x_k)$, is unity if $x = x_k$, and zero otherwise.

- Hence, for a discrete **RV**, the **expected** value is given by:

$$\mu_x = \int_{\mathbb{R}} x\, f_X\left(x\right)\, dx = \int_{\mathbb{R}} x \sum_k p_k\, \delta(x - x_k)\, dx = \sum_k x_k\, p_k$$

Properties of expectation operator

The expectation operator computes a statistical average by using the density $f_X(x)$ as a weighting function. Hence, the mean μ_x can be regarded as the *center of gravity* of the density.

Properties of expectation operator

The expectation operator computes a statistical average by using the density $f_X(x)$ as a weighting function. Hence, the mean μ_x can be regarded as the *center of gravity* of the density.

- If $f_X(x)$ is an even function, then $\mu_X = 0$. Note that since $f_X(x) \geq 0$, then $f_X(x)$ cannot be an odd function.
- If $f_X(x)$ is symmetrical about $x = a$, such that $f_X(a - x) = f_X(x + a)$, then $\mu_X = a$.

Properties of expectation operator

The expectation operator computes a statistical average by using the density $f_X(x)$ as a weighting function. Hence, the mean μ_x can be regarded as the *center of gravity* of the density.

- If $f_X(x)$ is an even function, then $\mu_X = 0$. Note that since $f_X(x) \geq 0$, then $f_X(x)$ cannot be an odd function.
- If $f_X(x)$ is symmetrical about $x = a$, such that $f_X(a - x) = f_X(x + a)$, then $\mu_X = a$.
- The expectation operator is linear:

$$\mathbb{E}\left[\alpha\, X(\zeta) + \beta\right] = \alpha\, \mu_X + \beta$$

Properties of expectation operator

- If $Y(\zeta) = g\{X(\zeta)\}$ is a **RV** obtained by transforming $X(\zeta)$ through a suitable function, the expectation of $Y(\zeta)$ is:

$$\mathbb{E}\left[Y(\zeta)\right] \triangleq \mathbb{E}\left[g\{X(\zeta)\}\right] = \int_{-\infty}^{\infty} g(x)\ f_X\left(x\right)\, dx$$

Properties of expectation operator

- If $Y(\zeta) = g\{X(\zeta)\}$ is a **RV** obtained by transforming $X(\zeta)$ through a suitable function, the expectation of $Y(\zeta)$ is:

$$\mathbb{E}\left[Y(\zeta)\right] \triangleq \mathbb{E}\left[g\{X(\zeta)\}\right] = \int_{-\infty}^{\infty} g(x)\ f_X\left(x\right)\ dx$$

  - This property means that you don't need to keep track of which pdf the expectation is taken with respect to.
  - Rather, you simply consider the RV inside the expectation, and the expectation is w. r. t. the pdf of that RV.

Properties of expectation operator

- If $Y(\zeta) = g\{X(\zeta)\}$ is a **RV** obtained by transforming $X(\zeta)$ through a suitable function, the expectation of $Y(\zeta)$ is:

$$\mathbb{E}\left[Y(\zeta)\right] \triangleq \mathbb{E}\left[g\{X(\zeta)\}\right] = \int_{-\infty}^{\infty} g(x)\; f_X(x)\; dx$$

  - This property means that you don't need to keep track of which pdf the expectation is taken with respect to.
  - Rather, you simply consider the RV inside the expectation, and the expectation is w. r. t. the pdf of that RV.
  - As an outline proof, consider a monotonic one-to-one function $y = g(x)$, such that $f_Y(y) = \frac{f_X(x)}{\frac{dy}{dx}}$.

Properties of expectation operator

- If $Y(\zeta) = g\{X(\zeta)\}$ is a **RV** obtained by transforming $X(\zeta)$ through a suitable function, the expectation of $Y(\zeta)$ is:

$$\mathbb{E}\left[Y(\zeta)\right] \triangleq \mathbb{E}\left[g\{X(\zeta)\}\right] = \int_{-\infty}^{\infty} g(x)\; f_X(x)\; dx$$

  - This property means that you don't need to keep track of which pdf the expectation is taken with respect to.
  - Rather, you simply consider the RV inside the expectation, and the expectation is w. r. t. the pdf of that RV.
  - As an outline proof, consider a monotonic one-to-one function $y = g(x)$, such that $f_Y(y) = \frac{f_X(x)}{\frac{dy}{dx}}$.

$$\mathbb{E}_{f_Y}\left[Y(\zeta)\right] = \int y f_Y(y)\; dy = \int g(x)\, \frac{f_X(x)}{\frac{dy}{dx}}\, dy = \int g(x)\, f_X(x)\; dx$$

Properties of expectation operator

Example (Trigonometric Transformation). The continuous random variable (RV), $\Theta(\zeta)$, is uniformally distributed between $-\pi$ and π.

1. Calculate the expected value of $\Theta(\zeta)$.
2. Now consider the RV, $Y(\zeta) = A\cos^2\Theta(\zeta)$, where A is assumed to be a constant value. What is the expected value of $Y(\zeta)$?

Properties of expectation operator

Example (Trigonometric Transformation). The continuous random variable (RV), $\Theta(\zeta)$, is uniformally distributed between $-\pi$ and π.

1. Calculate the expected value of $\Theta(\zeta)$.

2. Now consider the RV, $Y(\zeta) = A\,\cos^2\Theta(\zeta)$, where A is assumed to be a constant value. What is the expected value of $Y(\zeta)$?

SOLUTION. 1. The expected value of $\Theta(\zeta)$ is:

$$\mathbb{E}\left[\Theta(\zeta)\right] = \int_{-\infty}^{\infty} \theta\, f_\Theta(\theta)\, d\theta = \int_{-\pi}^{\pi} \theta\, \frac{1}{2\pi}\, d\theta$$

$$= \left.\frac{\theta^2}{4\pi}\right|_{-\pi}^{\pi} = 0 \qquad \square$$

Properties of expectation operator

Example (Trigonometric Transformation). The continuous random variable (RV), $\Theta(\zeta)$, is uniformally distributed between $-\pi$ and π.

1. Calculate the expected value of $\Theta(\zeta)$.
2. Now consider the RV, $Y(\zeta) = A\,\cos^2\Theta(\zeta)$, where A is assumed to be a constant value. What is the expected value of $Y(\zeta)$?

SOLUTION. 1. Using the invariance of the expectation operator:

$$\mathbb{E}[Y(\zeta)] = \mathbb{E}\left[A\cos^2\theta(\zeta)\right] = \int_{-\pi}^{\pi}\left[A\cos^2(\theta)\right] f_\Theta(\theta)\, d\theta$$

$$= \frac{A}{2\pi}\int_{-\pi}^{\pi}\cos^2(\theta)\, d\theta = \frac{A}{4\pi}\int_{-\pi}^{\pi}(1+\cos 2\theta)\, d\theta = \frac{A}{2} \quad \square$$

Properties of expectation operator

– End-of-Topic 21: **Expectations, their properties, and some examples** –

Any Questions?

Moments

Recall that **mean** and **variance** can be defined as:

$$\mathbb{E}\left[X(\zeta)\right] = \mu_X = \int_{\mathbb{R}} x\, f_X(x)\, dx$$

$$\text{var}\left[X(\zeta)\right] = \sigma_X^2 = \int_{\mathbb{R}} x^2\, f_X(x)\, dx - \mu_X^2 = \mathbb{E}\left[X^2(\zeta)\right] - \mathbb{E}^2\left[X(\zeta)\right]$$

Thus, key characteristics of the **pdf** of a **RV** can be calculated if the expressions $\mathbb{E}\left[X^m(\zeta)\right],\ m \in \{1, 2\}$ are known.

Moments

Recall that **mean** and **variance** can be defined as:

$$\mathbb{E}\left[X(\zeta)\right] = \mu_X = \int_{\mathbb{R}} x\, f_X(x)\, dx$$

$$\text{var}\left[X(\zeta)\right] = \sigma_X^2 = \int_{\mathbb{R}} x^2\, f_X(x)\, dx - \mu_X^2 = \mathbb{E}\left[X^2(\zeta)\right] - \mathbb{E}^2\left[X(\zeta)\right]$$

Thus, key characteristics of the **pdf** of a **RV** can be calculated if the expressions $\mathbb{E}\left[X^m(\zeta)\right],\ m \in \{1, 2\}$ are known.

Further aspects of the **pdf** can be described by defining various **moments** of $X(\zeta)$: the m-th moment of $X(\zeta)$ is given by:

$$r_X^{(m)} \triangleq \mathbb{E}\left[X^m(\zeta)\right] = \int_{\mathbb{R}} x^m\, f_X(x)\, dx$$

Note, of course, that in general: $\mathbb{E}\left[X^m(\zeta)\right] \neq \mathbb{E}^m\left[X(\zeta)\right]$.

Moments

Example (Exponential Random Variable). Calculate the moments of the exponential random variable with parameter λ. We can use:

$$\int_0^\infty u^n \, e^{-u} \, du = n! \qquad n \in \{0, 1, 2, \dots\}$$

$\bowtie$

Moments

Example (Exponential Random Variable). Calculate the moments of the exponential random variable with parameter λ. We can use:

$$\int_0^\infty u^n\, e^{-u}\, du = n! \qquad n \in \{0, 1, 2, \dots\}$$

SOLUTION. The pdf for an exponential RV is:

$$f_X(x) = \begin{cases} 0 & \text{if } x < 0, \\ \lambda e^{-\lambda x} & \text{if } x \geq 0, \end{cases}$$

□

Moments

Example (Exponential Random Variable). Calculate the moments of the exponential random variable with parameter λ. We can use:

$$\int_0^\infty u^n \, e^{-u} \, du = n! \qquad n \in \{0,\, 1,\, 2,\, \dots\}$$

SOLUTION. The pdf for an exponential RV is:

$$f_X(x) = \begin{cases} 0 & \text{if } x < 0, \\ \lambda e^{-\lambda x} & \text{if } x \geq 0, \end{cases}$$

The m-th moment is given by:

$$\mathbb{E}\left[X^m(\zeta)\right] = \int_0^\infty x^m \, f_X(x) \, dx = \lambda \int_0^\infty x^m \, e^{-\lambda x} \, dx$$

□

Moments

Example (Exponential Random Variable). Calculate the moments of the exponential random variable with parameter λ. We can use:

$$\int_0^\infty u^n \, e^{-u} \, du = n! \qquad n \in \{0, 1, 2, \dots\}$$

SOLUTION. The pdf for an exponential RV is:

$$f_X(x) = \begin{cases} 0 & \text{if } x < 0, \\ \lambda e^{-\lambda x} & \text{if } x \geq 0, \end{cases}$$

The m-th moment is given by:

$$\mathbb{E}\left[X^m(\zeta)\right] = \int_0^\infty x^m \, f_X(x) \, dx = \lambda \int_0^\infty x^m \, e^{-\lambda\, x} \, dx \qquad \square$$

Using the provided formula by setting $u = \lambda\, x$ such that when $x = \{0, \infty\}$ then $u = \{0, \infty\}$, and $du = \lambda \, dx$:

Moments

Example (Exponential Random Variable). Calculate the moments of the exponential random variable with parameter λ. We can use:

$$\int_0^\infty u^n e^{-u} \, du = n! \qquad n \in \{0, 1, 2, \dots\}$$

SOLUTION. Using the provided formula by setting $u = \lambda x$ such that when $x = \{0, \infty\}$ then $u = \{0, \infty\}$, and $du = \lambda \, dx$:

$$\mathbb{E}\left[X^m(\zeta)\right] = \frac{1}{\lambda^m} \int_0^\infty u^n e^{-u} \, du = \frac{m!}{\lambda^m}$$

□

Moments

Example (Exponential Random Variable). Calculate the moments of the exponential random variable with parameter λ. We can use:

$$\int_0^\infty u^n\, e^{-u}\, du = n! \qquad n \in \{0,\, 1,\, 2,\, \dots\}$$

SOLUTION. Using the provided formula by setting $u = \lambda\, x$ such that when $x = \{0,\, \infty\}$ then $u = \{0,\, \infty\}$, and $du = \lambda\, dx$:

$$\mathbb{E}\left[X^m(\zeta)\right] = \frac{1}{\lambda^m}\int_0^\infty u^n\, e^{-u}\, du = \frac{m!}{\lambda^m} \qquad \square$$

In particular, by setting $m = 1$, the mean is $\mu_X = \mathbb{E}\left[X(\zeta)\right] = {}^1/_\lambda$.

Setting $m = 2$, the second-moment is $\mathbb{E}\left[X^2(\zeta)\right] = {}^2/_{\lambda^2}$, and the variance is $\sigma_X^2 = \mathrm{var}\left[X(\zeta)\right] = {}^2/_{\lambda^2} - \left({}^1/_\lambda\right)^2 = \frac{1}{\lambda^2} = \mu_X^2$.

Moments

After this lecture, try the following example in the notes:

Example (Expectations of non-negative RVs). Let $X(\zeta)$ be a non-negative RV with pdf $f_X(x)$. Show that

$$\mathbb{E}[X^m(\zeta)] = \int_0^\infty m\, x^{m-1} \Pr(X(\zeta) > x)\, dx \qquad \bowtie$$

for any $m \geq 1$ for which the expectation is finite.

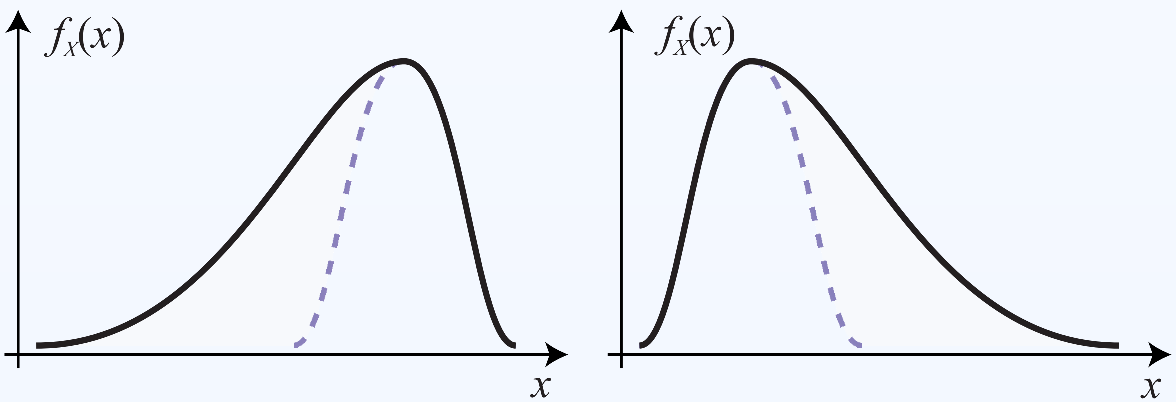
Higher-Order Statistics

Two important and commonly used higher-order statistics that are useful for characterising a random variable are:

Skewness characterises the degree of asymmetry of a distribution. It is a normalised third-order central moment:

$$\tilde{\kappa}_X^{(3)} \triangleq \mathbb{E}\left[\left\{\frac{X(\zeta)-\mu_X}{\sigma_X}\right\}^3\right] = \frac{1}{\sigma_X^3}\gamma_X^{(3)}$$

and is a *dimensionless* quantity.



Negative Skew　　Positive Skew

Higher-Order Statistics

Two important and commonly used higher-order statistics that are useful for characterising a random variable are:

Skewness characterises the degree of asymmetry of a distribution. It is a normalised third-order central moment:

$$\tilde{\kappa}_X^{(3)} \triangleq \mathbb{E}\left[\left\{\frac{X(\zeta) - \mu_X}{\sigma_X}\right\}^3\right] = \frac{1}{\sigma_X^3}\gamma_X^{(3)}$$

and is a *dimensionless* quantity.

- The **skewness** is:

$$\tilde{\kappa}_X^{(3)} = \begin{cases} < 0 & \text{if the density leans or stretches out towards the left} \\ 0 & \text{if the density is symmetric about } \mu_X \\ > 0 & \text{if the density leans or stretches out towards the right} \end{cases}$$

Higher-Order Statistics

Kurtosis measures relative flatness or *peakedness* of a distribution about its mean value.

Higher-Order Statistics

Kurtosis measures relative flatness or *peakedness* of a distribution about its mean value.

- It is defined based on a normalised fourth-central moment:

$$\tilde{\kappa}_X^{(4)} \triangleq \mathbb{E}\left[\left\{\frac{X(\zeta) - \mu_X}{\sigma_X}\right\}^4\right] - 3 = \frac{1}{\sigma_X^4}\gamma_X^{(4)} - 3$$

Higher-Order Statistics

Kurtosis measures relative flatness or *peakedness* of a distribution about its mean value.

- It is defined based on a normalised fourth-central moment:

$$\tilde{\kappa}_X^{(4)} \triangleq \mathbb{E}\left[\left\{\frac{X(\zeta)-\mu_X}{\sigma_X}\right\}^4\right] - 3 = \frac{1}{\sigma_X^4}\gamma_X^{(4)} - 3$$

This measure is relative with respect to a normal distribution, which has the property $\gamma_X^{(4)} = 3\sigma_X^4$, therefore having zero kurtosis.

Higher-Order Statistics

Example (Exponential distribution). Calculate the skewness of an exponential random variable with parameter λ.

Higher-Order Statistics

Example (Exponential distribution). Calculate the skewness of an exponential random variable with parameter λ.

SOLUTION. From earlier calculations it was was shown that the m-th moment was given by $r_X^{(m)} = {}^{m!}/\lambda^m$.

Higher-Order Statistics

Example (Exponential distribution). Calculate the skewness of an exponential random variable with parameter λ.

SOLUTION. From earlier calculations it was was shown that the m-th moment was given by $r_X^{(m)} = {m!}/{\lambda^m}$.

It can also be shown, by expanding the expression for skewness:

$$\tilde{\kappa}_X^{(3)} = \frac{r_X^{(3)} - 3 r_X^{(1)} r_X^{(2)} + 2(r_X^{(1)})^3}{\sigma_X^3}$$

□

Higher-Order Statistics

Example (Exponential distribution). Calculate the skewness of an exponential random variable with parameter λ.

SOLUTION. From earlier calculations it was was shown that the m-th moment was given by $r_X^{(m)} = {}^{m!}/_{\lambda^m}$.

It can also be shown, by expanding the expression for skewness:

$$\tilde{\kappa}_X^{(3)} = \frac{r_X^{(3)} - 3r_X^{(1)} r_X^{(2)} + 2(r_X^{(1)})^3}{\sigma_X^3}$$

Hence, since it was also shown that $\sigma_X^2 = {}^1/_{\lambda^2}$, then:

$$\tilde{\kappa}_X^{(3)} = \frac{\frac{3!}{\lambda^3} - 3\frac{1!}{\lambda}\frac{2!}{\lambda^2} + 2\frac{1}{\lambda^3}}{\frac{1}{\lambda^3}} = 2$$

□

Higher-Order Statistics

Example (Exponential distribution). Calculate the skewness of an exponential random variable with parameter λ.

SOLUTION. From earlier calculations it was was shown that the m-th moment was given by $r_X^{(m)} = {}^{m!}/_{\lambda^m}$.

It can also be shown, by expanding the expression for skewness:

$$\tilde{\kappa}_X^{(3)} = \frac{r_X^{(3)} - 3r_X^{(1)} r_X^{(2)} + 2(r_X^{(1)})^3}{\sigma_X^3}$$

Hence, since it was also shown that $\sigma_X^2 = {}^1/_{\lambda^2}$, then:

$$\tilde{\kappa}_X^{(3)} = \frac{\frac{3!}{\lambda^3} - 3\frac{1!}{\lambda}\frac{2!}{\lambda^2} + 2\frac{1}{\lambda^3}}{\frac{1}{\lambda^3}} = 2 \qquad \square$$

Positive skewness indicates leaning to the right, which it does!

Higher-Order Statistics

Example (Laplace distribution). Calculate the Kurtosis of the **standard Laplace distribution**, $f_X(x) = \frac{1}{2}e^{-|x|}$, $x \in \mathbb{R}$.

Higher-Order Statistics

Example (Laplace distribution). Calculate the Kurtosis of the **standard Laplace distribution**, $f_X(x) = \frac{1}{2}e^{-|x|}$, $x \in \mathbb{R}$.

SOLUTION. As the density is symmetric, the skewness is zero! Moreover, the odd moments are also equal to zero through symmetry (left as an exercise).

Higher-Order Statistics

Example (Laplace distribution). Calculate the Kurtosis of the **standard Laplace distribution**, $f_X(x) = \frac{1}{2} e^{-|x|}$, $x \in \mathbb{R}$.

SOLUTION. As the density is symmetric, the skewness is zero! Moreover, the odd moments are also equal to zero through symmetry (left as an exercise).

The even moments are given by:

$$r_X^{(m)} = \frac{1}{2} \int_{-\infty}^{0} x^m \, e^x \, dx + \frac{1}{2} \int_{0}^{\infty} x^m \, e^{-x} \, dx = \int_{0}^{\infty} x^m \, e^{-x} \, dx = m!$$

□

Higher-Order Statistics

Example (Laplace distribution). Calculate the Kurtosis of the **standard Laplace distribution**, $f_X(x) = \frac{1}{2} e^{-|x|}$, $x \in \mathbb{R}$.

SOLUTION. As the density is symmetric, the skewness is zero! Moreover, the odd moments are also equal to zero through symmetry (left as an exercise).

The even moments are given by:

$$r_X^{(m)} = \frac{1}{2} \int_{-\infty}^{0} x^m \, e^x \, dx + \frac{1}{2} \int_{0}^{\infty} x^m \, e^{-x} \, dx = \int_{0}^{\infty} x^m \, e^{-x} \, dx = m!$$

Hence, using the formula for Kurtosis (noting $r_X^{(1)} = 0$):

$$\tilde{\kappa}_X^{(4)} = \mathbb{E}\left[\left\{ \frac{X(\zeta) - \mu_X}{\sigma_X} \right\}^4 \right] - 3 = \frac{r_X^{(4)}}{\left(r_X^{(2)}\right)^2} - 3 = \frac{4!}{(2!)^2} - 3 = 3 \quad \square$$

Higher-Order Statistics

Skewness and kurtosis are used in signal processing in the following applications:

Higher-Order Statistics

Skewness and kurtosis are used in signal processing in the following applications:

Signal Separation is only possible if the signals are statistically distinctive and this requires non-Gaussianity; maximising kurtosis means that separated signals are ensured to be as non-Gaussian as possible.

Higher-Order Statistics

Skewness and kurtosis are used in signal processing in the following applications:

Signal Separation is only possible if the signals are statistically distinctive and this requires non-Gaussianity; maximising kurtosis means that separated signals are ensured to be as non-Gaussian as possible.

Outlier detection As kurtosis is a measure of heaviness of the tails, it also provides a metric for the number of outliers. Outliers, for example positive values, can also lead to asymmetric densities, measured by skewness.

Higher-Order Statistics

Skewness and kurtosis are used in signal processing in the following applications:

Signal Separation is only possible if the signals are statistically distinctive and this requires non-Gaussianity; maximising kurtosis means that separated signals are ensured to be as non-Gaussian as possible.

Outlier detection As kurtosis is a measure of heaviness of the tails, it also provides a metric for the number of outliers. Outliers, for example positive values, can also lead to asymmetric densities, measured by skewness.

Features Skewness and kurtosis can be used in feature-based classification and machine learning algorithms.

Higher-Order Statistics

– End-of-Topic 22: **Skewness, Kurtosis, and their Applications** –

Any Questions?

Lecture Slideset 3
Multiple Random Variables

Abstract

A *group* of signal observations can be modelled as a collection of random variables (RVs) that can be grouped to form a **random vector**, or **vector RV**.

- This is an extension of the concept of a RV, and generalises many of the results presented for scalar RVs.

Abstract

A *group* of signal observations can be modelled as a collection of random variables (RVs) that can be grouped to form a **random vector**, or **vector RV**.

- This is an extension of the concept of a RV, and generalises many of the results presented for scalar RVs.
- Note that each element of a **random vector** is not necessarily generated independently from a separate *experiment*.

Abstract

A *group* of signal observations can be modelled as a collection of random variables (RVs) that can be grouped to form a **random vector**, or **vector RV**.

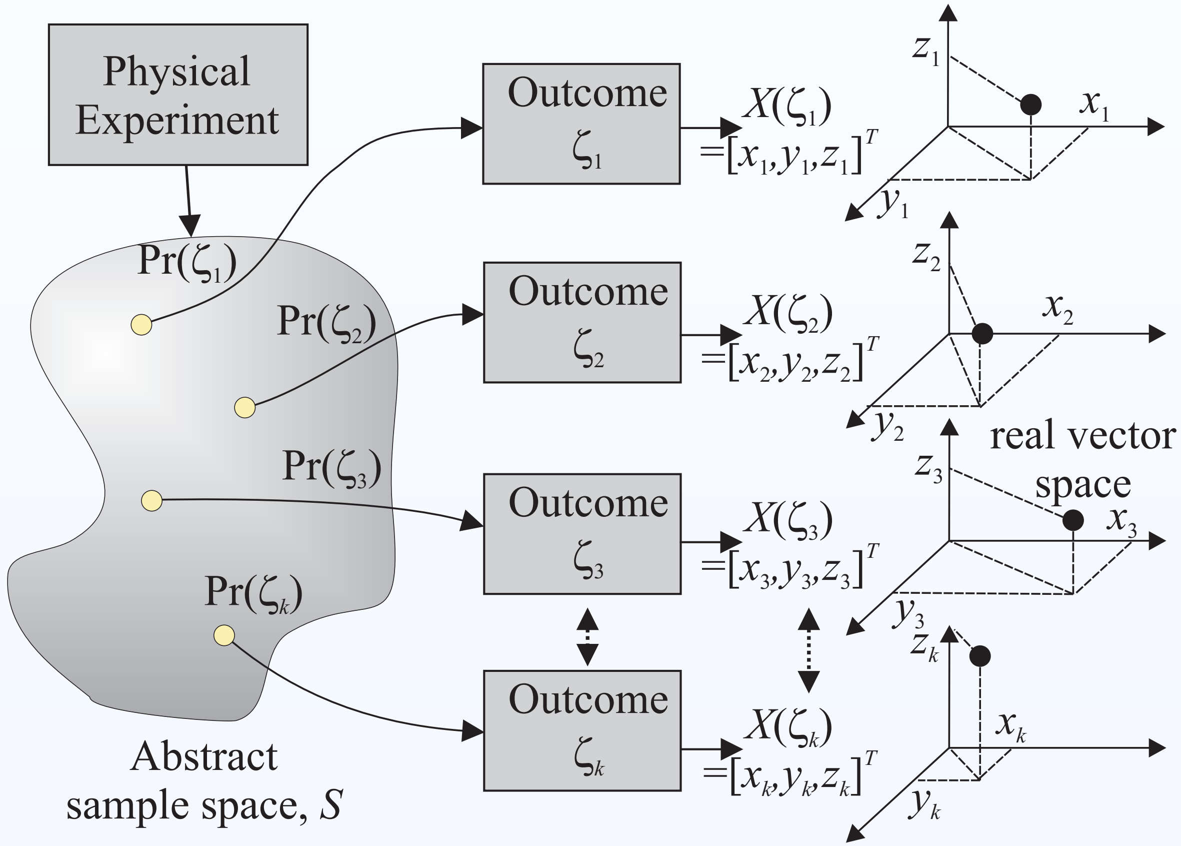
- This is an extension of the concept of a RV, and generalises many of the results presented for scalar RVs.
- Note that each element of a **random vector** is not necessarily generated independently from a separate *experiment*.
- Random vectors also lead to the notion of the relationship between the random elements.

Abstract

A *group* of signal observations can be modelled as a collection of random variables (RVs) that can be grouped to form a **random vector**, or **vector RV**.

- This is an extension of the concept of a RV, and generalises many of the results presented for scalar RVs.
- Note that each element of a **random vector** is not necessarily generated independently from a separate *experiment*.
- Random vectors also lead to the notion of the relationship between the random elements.
- This course mainly deals with real-valued random vectors, although the concept can be extended to complex-valued random vectors.

Definition of Random Vectors



A graphical representation of a random vector.

Definition of Random Vectors

A real-valued random vector $\mathbf{X}(\zeta)$ containing N real-valued RVs, each denoted by $X_n(\zeta)$ for $n \in \mathcal{N} = \{1, \dots, N\}$, is denoted by the column-vector:

$$\mathbf{X}(\zeta) = \begin{bmatrix} X_1(\zeta) & X_2(\zeta) & \cdots & X_N(\zeta) \end{bmatrix}^T$$

Definition of Random Vectors

A real-valued random vector $\mathbf{X}(\zeta)$ containing N real-valued RVs, each denoted by $X_n(\zeta)$ for $n \in \mathcal{N} = \{1, \dots, N\}$, is denoted by the column-vector:

$$\mathbf{X}(\zeta) = \begin{bmatrix} X_1(\zeta) & X_2(\zeta) & \cdots & X_N(\zeta) \end{bmatrix}^T$$

A real-valued random vector can be thought as a mapping from an abstract probability space to a vector-valued, real space $\mathbb{R}^N$.

Definition of Random Vectors

A real-valued random vector $\mathbf{X}(\zeta)$ containing N real-valued RVs, each denoted by $X_n(\zeta)$ for $n \in \mathcal{N} = \{1, \dots, N\}$, is denoted by the column-vector:

$$\mathbf{X}(\zeta) = \begin{bmatrix} X_1(\zeta) & X_2(\zeta) & \cdots & X_N(\zeta) \end{bmatrix}^T$$

A real-valued random vector can be thought as a mapping from an abstract probability space to a vector-valued, real space $\mathbb{R}^N$.

Denote a specific value for a random vector as:

$$\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_N \end{bmatrix}^T$$

Then the notation $\mathbf{X}(\zeta) \leq \mathbf{x}$ is equivalent to the event $\{X_n(\zeta) \leq x_n, \, n \in \mathcal{N}\}$.

Distribution and Density Functions

The **joint cdf** completely characterises a random vector, and is defined by:

$$F_{\mathbf{X}}\left(\mathbf{x}\right) \triangleq \Pr\left(\{X_n(\zeta) \leq x_n,\, n \in \mathcal{N}\}\right) = \Pr\left(\mathbf{X}\left(\zeta\right) \leq \mathbf{x}\right)$$

Distribution and Density Functions

The **joint cdf** completely characterises a random vector, and is defined by:

$$F_{\mathbf{X}}(\mathbf{x}) \triangleq \Pr\left(\{X_n(\zeta) \le x_n,\, n \in \mathcal{N}\}\right) = \Pr\left(\mathbf{X}(\zeta) \le \mathbf{x}\right)$$

A random vector can also be characterised by its **joint pdf**, which is defined by:

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= \lim_{\Delta\mathbf{x}\to\mathbf{0}} \frac{\Pr\left(\{x_n < X_n(\zeta) \le x_n + \Delta x_n,\, n \in \mathcal{N}\}\right)}{\Delta x_1 \cdots \Delta x_N} \\ &= \frac{\partial}{\partial x_1}\frac{\partial}{\partial x_2}\cdots\frac{\partial}{\partial x_N} F_{\mathbf{X}}(\mathbf{x}) \end{aligned}$$

Distribution and Density Functions

The **joint cdf** completely characterises a random vector, and is defined by:

$$F_{\mathbf{X}}(\mathbf{x}) \triangleq \Pr\left(\{X_n(\zeta) \le x_n,\, n \in \mathcal{N}\}\right) = \Pr\left(\mathbf{X}(\zeta) \le \mathbf{x}\right)$$

A random vector can also be characterised by its **joint pdf**, which is defined by:

$$f_{\mathbf{X}}(\mathbf{x}) = \lim_{\Delta\mathbf{x}\to\mathbf{0}} \frac{\Pr\left(\{x_n < X_n(\zeta) \le x_n + \Delta x_n,\, n \in \mathcal{N}\}\right)}{\Delta x_1 \cdots \Delta x_N}$$
$$= \frac{\partial}{\partial x_1}\frac{\partial}{\partial x_2}\cdots\frac{\partial}{\partial x_N} F_{\mathbf{X}}(\mathbf{x})$$

Hence, it follows:

$$F_{\mathbf{X}}(\mathbf{x}) = \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_N} f_{\mathbf{X}}(\mathbf{v})\, dv_N \cdots dv_1 = \int_{-\infty}^{\mathbf{x}} f_{\mathbf{X}}(\mathbf{v})\, d\mathbf{v}$$

Distribution and Density Functions

– End-of-Topic 23: **Introduction to Random Vectors, its definition, and joint distribution and density functions** –

Any Questions?

Distribution and Density Functions

- Properties of **joint-cdf**:

$$0 \leq F_{\mathbf{X}}(\mathbf{x}) \leq 1, \quad \lim_{\mathbf{x} \to -\infty} F_{\mathbf{X}}(\mathbf{x}) = 0, \quad \lim_{\mathbf{x} \to \infty} F_{\mathbf{X}}(\mathbf{x}) = 1$$

$F_{\mathbf{X}}(\mathbf{x})$ is a monotonically increasing function of $\mathbf{x}$:

$$F_{\mathbf{X}}(\mathbf{a}) \leq F_{\mathbf{X}}(\mathbf{b}) \quad \text{if} \quad \mathbf{a} \leq \mathbf{b}$$

Distribution and Density Functions

- Properties of **joint-cdf**:

$$0 \leq F_{\mathbf{X}}(\mathbf{x}) \leq 1, \quad \lim_{\mathbf{x} \to -\infty} F_{\mathbf{X}}(\mathbf{x}) = 0, \quad \lim_{\mathbf{x} \to \infty} F_{\mathbf{X}}(\mathbf{x}) = 1$$

$F_{\mathbf{X}}(\mathbf{x})$ is a monotonically increasing function of $\mathbf{x}$:

$$F_{\mathbf{X}}(\mathbf{a}) \leq F_{\mathbf{X}}(\mathbf{b}) \quad \text{if} \quad \mathbf{a} \leq \mathbf{b}$$

- Properties of **joint-pdfs**:

$$f_{\mathbf{X}}(\mathbf{x}) \geq 0, \quad \int_{-\infty}^{\infty} f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x} = 1$$

Distribution and Density Functions

- Properties of **joint-cdf**:

$$0 \leq F_{\mathbf{X}}(\mathbf{x}) \leq 1, \quad \lim_{\mathbf{x}\to-\infty} F_{\mathbf{X}}(\mathbf{x}) = 0, \quad \lim_{\mathbf{x}\to\infty} F_{\mathbf{X}}(\mathbf{x}) = 1$$

$F_{\mathbf{X}}(\mathbf{x})$ is a monotonically increasing function of $\mathbf{x}$:

$$F_{\mathbf{X}}(\mathbf{a}) \leq F_{\mathbf{X}}(\mathbf{b}) \quad \text{if} \quad \mathbf{a} \leq \mathbf{b}$$

- Properties of **joint-pdfs**:

$$f_{\mathbf{X}}(\mathbf{x}) \geq 0, \quad \int_{-\infty}^{\infty} f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x} = 1$$

- Probability of arbitrary events; note that

$$\Pr(\mathbf{x}_1 < \mathbf{X}(\zeta) \leq \mathbf{x}_2) = \int_{\mathbf{x}_1}^{\mathbf{x}_2} f_{\mathbf{X}}(\mathbf{v})\, d\mathbf{v} \neq F_{\mathbf{X}}(\mathbf{x}_2) - F_{\mathbf{X}}(\mathbf{x}_1)$$

Distribution and Density Functions

Example ([Therrien:1992, Example 2.1, Page 20]). The joint-pdf of a random vector $\mathbf{Z}(\zeta)$ which has two elements and therefore two random variables given by $X(\zeta)$ and $Y(\zeta)$ is given by:

$$f_{\mathbf{Z}}(\mathbf{z}) = \begin{cases} \frac{1}{2}(x+3y) & 0 \leq \{x,\, y\} \leq 1 \\ 0 & \text{otherwise} \end{cases} \qquad \bowtie$$

Calculate the joint-cumulative distribution function, $F_{\mathbf{Z}}(\mathbf{z})$.

Distribution and Density Functions

Example ([Therrien:1992, Example 2.1, Page 20]). The joint-pdf of a random vector $\mathbf{Z}(\zeta)$ which has two elements and therefore two random variables given by $X(\zeta)$ and $Y(\zeta)$ is given by:

$$f_{\mathbf{Z}}(\mathbf{z}) = \begin{cases} \frac{1}{2}(x+3y) & 0 \le \{x,\, y\} \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Calculate the joint-cumulative distribution function, $F_{\mathbf{Z}}(\mathbf{z})$.

SOLUTION. First note that the pdf integrates to unity since:

$$\int_{-\infty}^{\infty} f_{\mathbf{Z}}(\mathbf{z})\, d\mathbf{z} = \int_0^1 \int_0^1 \frac{1}{2}(x+3y)\, dx\, dy = \int_0^1 \frac{1}{2}\left[\frac{1}{2}x^2 + 3xy\right]_0^1 dy$$

□

Distribution and Density Functions

Example ([Therrien:1992, Example 2.1, Page 20]).

$$f_{\mathbf{Z}}(\mathbf{z}) = \begin{cases} \frac{1}{2}(x+3y) & 0 \le \{x,\, y\} \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Calculate the joint-cumulative distribution function, $F_{\mathbf{Z}}(\mathbf{z})$.

SOLUTION. First note that the pdf integrates to unity since:

$$\int_{-\infty}^{\infty} f_{\mathbf{Z}}(\mathbf{z})\, d\mathbf{z} = \int_0^1 \int_0^1 \frac{1}{2}(x+3y)\, dx\, dy = \int_0^1 \frac{1}{2}\left[\frac{1}{2}x^2 + 3xy\right]_0^1 dy$$

$$= \int_0^1 \frac{1}{4} + \frac{3}{2} y\, dy = \left[\frac{y}{4} + \frac{3y^2}{4}\right]_0^1 = \frac{1}{4} + \frac{3}{4} = 1$$

□

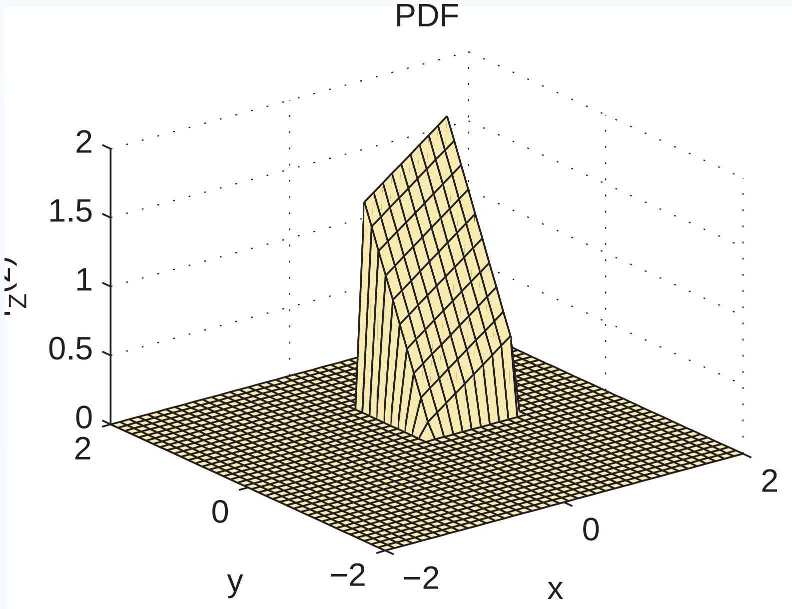
Distribution and Density Functions

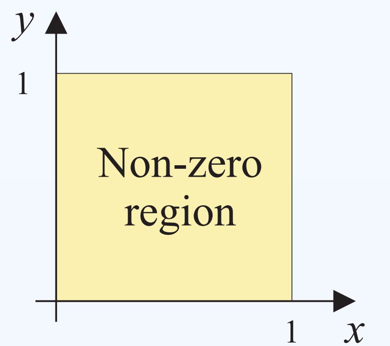
Example ([Therrien:1992, Example 2.1, Page 20]).

$$f_{\mathbf{Z}}(\mathbf{z}) = \begin{cases} \frac{1}{2}(x+3y) & 0 \le \{x,\, y\} \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Calculate the joint-cumulative distribution function, $F_{\mathbf{Z}}(\mathbf{z})$.

SOLUTION. The pdf is shown here:





Distribution and Density Functions

Example ([Therrien:1992, Example 2.1, Page 20]).

$$f_{\mathbf{Z}}(\mathbf{z}) = \begin{cases} \frac{1}{2}(x+3y) & 0 \le \{x,\, y\} \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Calculate the joint-cumulative distribution function, $F_{\mathbf{Z}}(\mathbf{z})$.

SOLUTION. For $x \le 0$ or $y \le 0$, $f_{\mathbf{Z}}(\mathbf{z}) = 0$, and thus $F_{\mathbf{Z}}(\mathbf{z}) = 0$.

Distribution and Density Functions

Example ([Therrien:1992, Example 2.1, Page 20]).

$$f_{\mathbf{Z}}(\mathbf{z}) = \begin{cases} \frac{1}{2}(x+3y) & 0 \le \{x,\, y\} \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Calculate the joint-cumulative distribution function, $F_{\mathbf{Z}}(\mathbf{z})$.

SOLUTION. For $x \le 0$ or $y \le 0$, $f_{\mathbf{Z}}(\mathbf{z}) = 0$, and thus $F_{\mathbf{Z}}(\mathbf{z}) = 0$.

If $0 < x \le 1$ *and* $0 < y \le 1$, the cdf is given by:

$$F_{\mathbf{Z}}(\mathbf{z}) = \int_{-\infty}^{\mathbf{z}} f_{\mathbf{Z}}(\bar{\mathbf{z}})\, d\bar{\mathbf{z}} = \int_0^y \int_0^x \frac{1}{2}(\bar{x} + 3\bar{y})\, d\bar{x}\, d\bar{y}$$

□

Distribution and Density Functions

Example ([Therrien:1992, Example 2.1, Page 20]).

$$f_{\mathbf{Z}}\left(\mathbf{z}\right) = \begin{cases} \frac{1}{2}(x+3y) & 0 \leq \{x,\, y\} \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Calculate the joint-cumulative distribution function, $F_{\mathbf{Z}}\left(\mathbf{z}\right)$.

SOLUTION. For $x \leq 0$ or $y \leq 0$, $f_{\mathbf{Z}}\left(\mathbf{z}\right) = 0$, and thus $F_{\mathbf{Z}}\left(\mathbf{z}\right) = 0$.

If $0 < x \leq 1$ *and* $0 < y \leq 1$, the cdf is given by:

$$F_{\mathbf{Z}}\left(\mathbf{z}\right) = \int_{-\infty}^{\mathbf{z}} f_{\mathbf{Z}}\left(\bar{\mathbf{z}}\right)\, d\bar{\mathbf{z}} = \int_0^y \int_0^x \frac{1}{2}\left(\bar{x}+3\bar{y}\right)\, d\bar{x}\, d\bar{y}$$

$$= \int_0^y \frac{1}{2}\left(\frac{x^2}{2} + 3x\bar{y}\right)\, d\bar{y} = \frac{1}{2}\left(\frac{x^2}{2}y + \frac{3xy^2}{2}\right) = \frac{xy}{4}(x+3y)$$

□

Distribution and Density Functions

Example ([Therrien:1992, Example 2.1, Page 20]).

$$f_{\mathbf{Z}}(\mathbf{z}) = \begin{cases} \frac{1}{2}(x+3y) & 0 \le \{x, y\} \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Calculate the joint-cumulative distribution function, $F_{\mathbf{Z}}(\mathbf{z})$.

SOLUTION. For $x \le 0$ or $y \le 0$, $f_{\mathbf{Z}}(\mathbf{z}) = 0$, and thus $F_{\mathbf{Z}}(\mathbf{z}) = 0$.

If $0 < x \le 1$ *and* $0 < y \le 1$, the cdf is given by:

$$F_{\mathbf{Z}}(\mathbf{z}) = \int_{-\infty}^{\mathbf{z}} f_{\mathbf{Z}}(\bar{\mathbf{z}})\, d\bar{\mathbf{z}} = \int_0^y \int_0^x \frac{1}{2}(\bar{x} + 3\bar{y})\, d\bar{x}\, d\bar{y}$$

$$= \int_0^y \frac{1}{2}\left(\frac{x^2}{2} + 3x\bar{y}\right) d\bar{y} = \frac{1}{2}\left(\frac{x^2}{2}y + \frac{3xy^2}{2}\right) = \frac{xy}{4}(x+3y) \quad \square$$

Finally, if $x > 1$ or $y > 1$, the upper limit becomes equal to 1.

Distribution and Density Functions

Example ([Therrien:1992, Example 2.1, Page 20]).

$$f_{\mathbf{Z}}(\mathbf{z}) = \begin{cases} \frac{1}{2}(x+3y) & 0 \le \{x, y\} \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Calculate the joint-cumulative distribution function, $F_{\mathbf{Z}}(\mathbf{z})$.

SOLUTION. Hence, in summary, it follows:

$$F_{\mathbf{Z}}(\mathbf{z}) = \begin{cases} 0 & x \le 0 \quad \text{or} \quad y \le 0 \\ \frac{xy}{4}(x+3y) & 0 < x, y \le 1 \\ \frac{x}{4}(x+3) & 0 < x \le 1, 1 < y \\ \frac{y}{4}(1+3y) & 0 < y \le 1, 1 < x \\ 1 & 1 < x, y < \infty \end{cases}$$

□

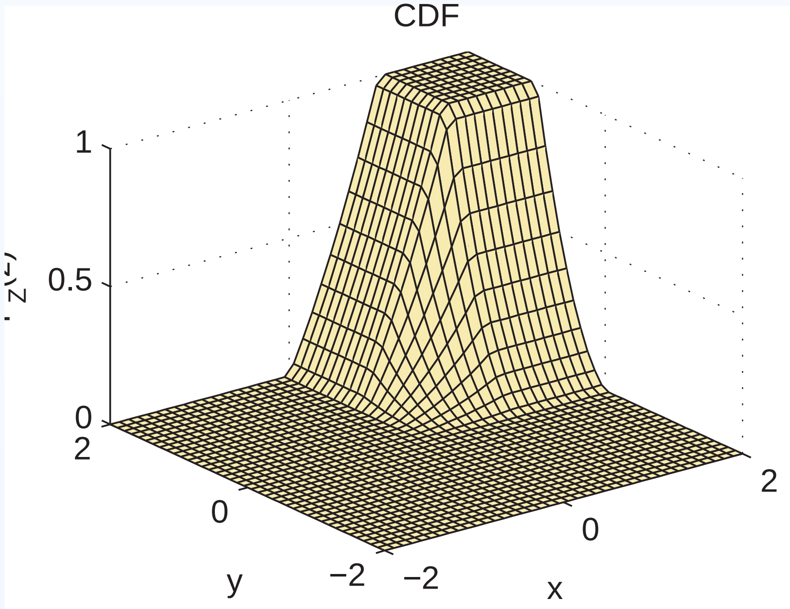
Distribution and Density Functions

Example ([Therrien:1992, Example 2.1, Page 20]).

$$f_{\mathbf{Z}}(\mathbf{z}) = \begin{cases} \frac{1}{2}(x+3y) & 0 \le \{x,\, y\} \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Calculate the joint-cumulative distribution function, $F_{\mathbf{Z}}(\mathbf{z})$.

SOLUTION. The cdf is plotted here:



A plot of the cumulative distribution function.

Distribution and Density Functions

– End-of-Topic 24: **Properties and Examples of Joint Distributions and Densities** –

Any Questions?

Marginal Density Function

The joint pdf characterises the random vector; the so-called **marginal pdf** describes a subset of RVs from the random vector.

Marginal Density Function

The joint pdf characterises the random vector; the so-called **marginal pdf** describes a subset of RVs from the random vector.

Let $\mathbf{k}$ be an M-dimensional vector containing unique indices to elements in the N-dimensional random vector $\mathbf{X}(\zeta)$,

$$\mathbf{k} = \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_M \end{bmatrix}$$

Marginal Density Function

The joint pdf characterises the random vector; the so-called **marginal pdf** describes a subset of RVs from the random vector.

Let $\mathbf{k}$ be an M-dimensional vector containing unique indices to elements in the N-dimensional random vector $\mathbf{X}(\zeta)$,

Now define a M-dimensional random vector, $\mathbf{X}_{\mathbf{k}}(\zeta)$, that contains the M random variables which are components of $\mathbf{X}(\zeta)$ and indexed by the elements of $\mathbf{k}$. In other-words, if

$$\mathbf{k} = \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_M \end{bmatrix} \quad \text{then} \quad \mathbf{X}_{\mathbf{k}}(\zeta) = \begin{bmatrix} X_{k_1}(\zeta) \\ X_{k_2}(\zeta) \\ \vdots \\ X_{k_M}(\zeta) \end{bmatrix}$$

Marginal Density Function

The **marginal pdf** is then given by:

$$f_{\mathbf{X_k}}(\mathbf{x_k}) = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{N-M \text{ integrals}} f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x_{-k}}$$

Marginal Density Function

The **marginal pdf** is then given by:

$$f_{\mathbf{x_k}}(\mathbf{x_k}) = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{N-M \text{ integrals}} f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x_{-k}}$$

A special case is the **marginal pdf** describing the individual RV X_j:

$$f_{X_j}(x_j) = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{N-1 \text{ integrals}} f_{\mathbf{X}}(\mathbf{x})\, dx_1 \cdots dx_{j-1} dx_{j+1} \cdots dx_N$$

Marginal Density Function

The **marginal pdf** is then given by:

$$f_{\mathbf{X_k}}(\mathbf{x_k}) = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{N-M \text{ integrals}} f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}_{-\mathbf{k}}$$

A special case is the **marginal pdf** describing the individual RV X_j:

$$f_{X_j}(x_j) = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{N-1 \text{ integrals}} f_{\mathbf{X}}(\mathbf{x})\, dx_1 \cdots dx_{j-1} dx_{j+1} \cdots dx_N$$

Marginal pdfs will become particular useful when dealing with Bayesian parameter estimation later in the course.

Marginal Density Function

Example (Marginalisation). The joint-pdf of a random vector $\mathbf{Z}(\zeta)$ which has two elements and therefore two random variables given by $X(\zeta)$ and $Y(\zeta)$ is given by:

$$f_{\mathbf{Z}}(\mathbf{z}) = \begin{cases} \frac{1}{2}(x+3y) & 0 \leq \{x,\, y\} \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

⋈

Calculate the marginal-pdfs, $f_X(x)$ and $f_Y(y)$, and their corresponding marginal-cdfs, $F_X(x)$ and $F_Y(y)$.

Marginal Density Function

Example (Marginalisation). The joint-pdf of a random vector $\mathbf{Z}(\zeta)$ which has two elements and therefore two random variables given by $X(\zeta)$ and $Y(\zeta)$ is given by:

$$f_{\mathbf{Z}}(\mathbf{z}) = \begin{cases} \frac{1}{2}(x+3y) & 0 \le \{x, y\} \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Calculate the marginal-pdfs, $f_X(x)$ and $f_Y(y)$, and their corresponding marginal-cdfs, $F_X(x)$ and $F_Y(y)$.

SOLUTION. By definition:

$$f_X(x) = \int_{\mathbb{R}} f_{\mathbf{Z}}(\mathbf{z})\, dy$$

$$f_Y(y) = \int_{\mathbb{R}} f_{\mathbf{Z}}(\mathbf{z})\, dx$$

Marginal Density Function

Example (Marginalisation).

$$f_{\mathbf{Z}}(\mathbf{z}) = \begin{cases} \frac{1}{2}(x+3y) & 0 \leq \{x, y\} \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Calculate the marginal-pdfs, $f_X(x)$ and $f_Y(y)$, and their corresponding marginal-cdfs, $F_X(x)$ and $F_Y(y)$.

SOLUTION. Taking $f_X(x)$, then:

$$f_X(x) = \begin{cases} \frac{1}{2}\int_0^1 (x+3y)\,dy & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

□

Marginal Density Function

Example (Marginalisation).

$$f_{\mathbf{Z}}(\mathbf{z}) = \begin{cases} \frac{1}{2}(x+3y) & 0 \leq \{x, y\} \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Calculate the marginal-pdfs, $f_X(x)$ and $f_Y(y)$, and their corresponding marginal-cdfs, $F_X(x)$ and $F_Y(y)$.

SOLUTION. Taking $f_X(x)$, then:

$$f_X(x) = \begin{cases} \frac{1}{2}\int_0^1 (x+3y)\,dy & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

which after a simple integration gives:

$$f_X(x) = \begin{cases} \frac{1}{2}\left(x+\frac{3}{2}\right) & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Marginal Density Function

Example (Marginalisation).

$$f_{\mathbf{Z}}(\mathbf{z}) = \begin{cases} \frac{1}{2}(x+3y) & 0 \le \{x, y\} \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Calculate the marginal-pdfs, $f_X(x)$ and $f_Y(y)$, and their corresponding marginal-cdfs, $F_X(x)$ and $F_Y(y)$.

SOLUTION. The cdf, $F_X(x)$, is thus given by:

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\, du = \begin{cases} 0 & x \le 0 \\ \frac{1}{2}\int_0^x \left(u + \frac{3}{2}\right) du & 0 \le x \le 1 \\ \frac{1}{2}\int_0^1 \left(u + \frac{3}{2}\right) du & x > 1 \end{cases}$$

□

Marginal Density Function

Example (Marginalisation).

$$f_{\mathbf{Z}}(\mathbf{z}) = \begin{cases} \frac{1}{2}(x+3y) & 0 \le \{x, y\} \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Calculate the marginal-pdfs, $f_X(x)$ and $f_Y(y)$, and their corresponding marginal-cdfs, $F_X(x)$ and $F_Y(y)$.

SOLUTION. The cdf, $F_X(x)$, is thus given by:

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\, du = \begin{cases} 0 & x \le 0 \\ \frac{1}{2}\int_0^x \left(u + \frac{3}{2}\right) du & 0 \le x \le 1 \\ \frac{1}{2}\int_0^1 \left(u + \frac{3}{2}\right) du & x > 1 \end{cases}$$

$$F_X(x) = \begin{cases} 0 & x \le 0 \\ \frac{x}{4}(x+3) & 0 \le x \le 1 \\ 1 & x > 1 \end{cases}$$

Marginal Density Function

Example (Marginalisation).

$$f_{\mathbf{Z}}(\mathbf{z}) = \begin{cases} \frac{1}{2}(x+3y) & 0 \leq \{x, y\} \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Calculate the marginal-pdfs, $f_X(x)$ and $f_Y(y)$, and their corresponding marginal-cdfs, $F_X(x)$ and $F_Y(y)$.

SOLUTION. Similarly, it can be shown that:

$$f_Y(y) = \begin{cases} \frac{1}{2}\left(\frac{1}{2}+3y\right) & 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

and

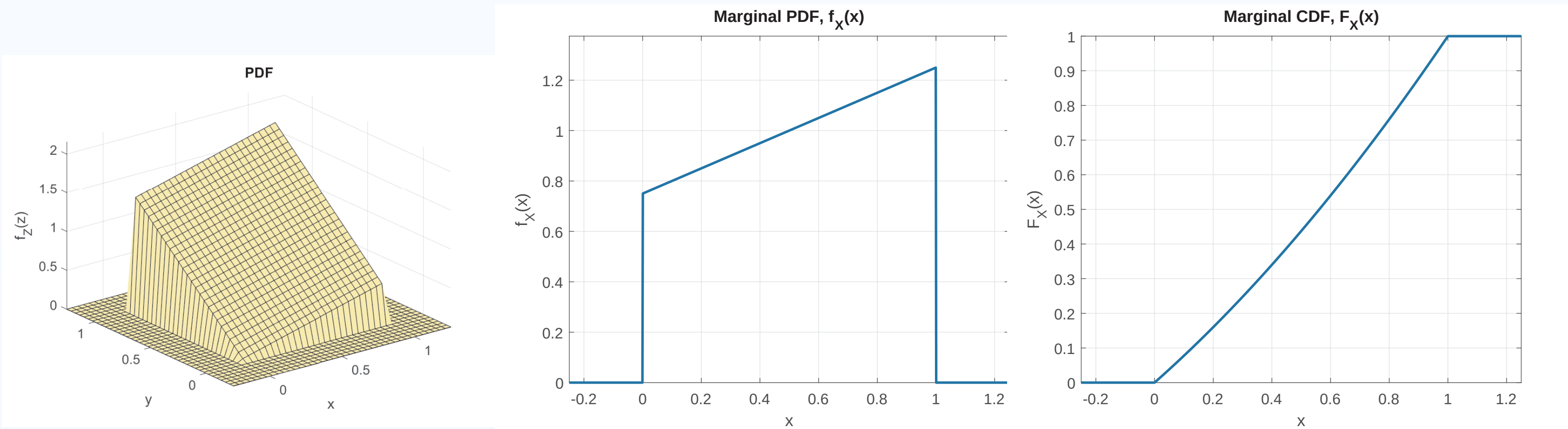
$$F_Y(y) = \begin{cases} 0 & y \leq 0 \\ \frac{y}{4}(1+3y) & 0 \leq y \leq 1 \\ 1 & y > 1 \end{cases}$$

Marginal Density Function

Example (Marginalisation).

$$f_{\mathbf{Z}}(\mathbf{z}) = \begin{cases} \frac{1}{2}(x+3y) & 0 \leq \{x, y\} \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

SOLUTION. The marginal-pdfs and cdfs are shown below.



The marginal-pdf, $f_X(x)$, and cdf, $F_X(x)$, for the RV, $X(\zeta)$.

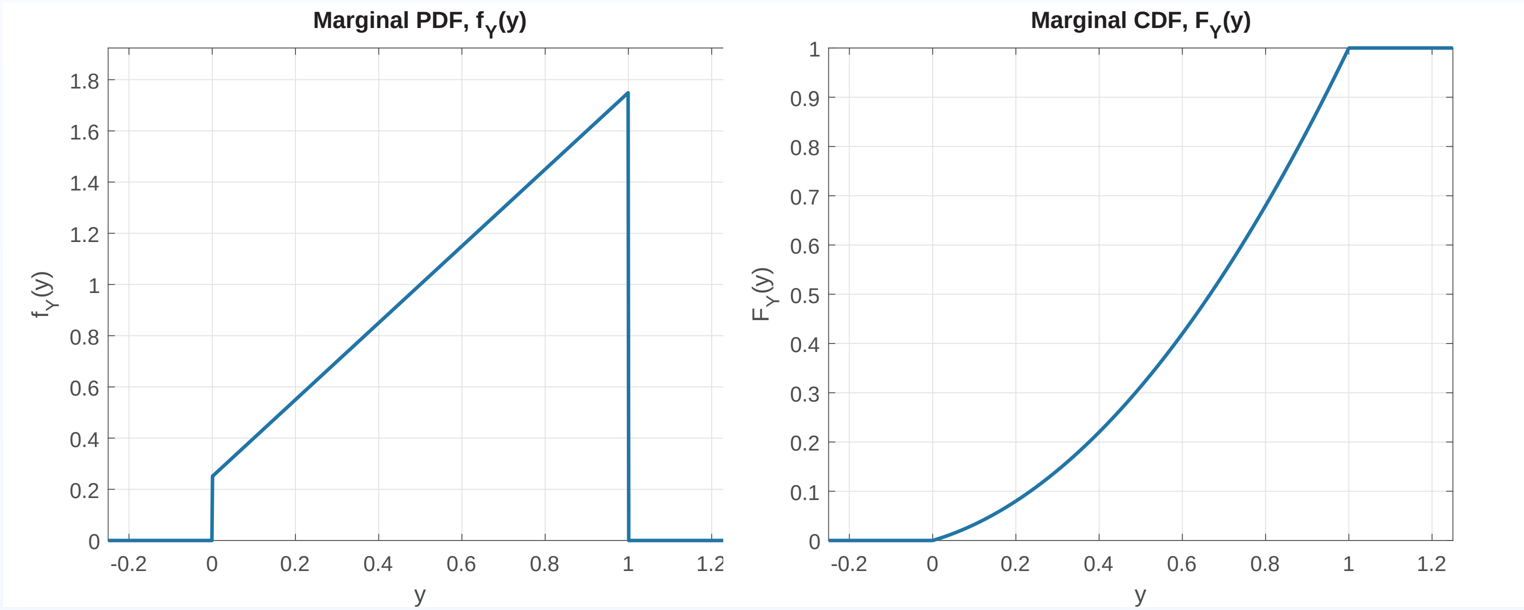
- Note that the marginal-pdf is not a *slice* of the joint-pdf.
- It is the integral of the joint-pdf over the other variable along a line whose position corresponds to the value of interest. □

Marginal Density Function

Example (Marginalisation).

$$f_{\mathbf{Z}}(\mathbf{z}) = \begin{cases} \frac{1}{2}(x+3y) & 0 \leq \{x, y\} \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

SOLUTION. The marginal-pdfs and cdfs are shown below.



The marginal-pdf, $f_Y(y)$, and cdf, $F_Y(y)$, for the RV, $Y(\zeta)$.

Marginal Density Function

– End-of-Topic 25: **Marginal Densities and Distributions and their Applications** –

Any Questions?

Independence

Two random variables, $X_1(\zeta)$ and $X_2(\zeta)$ are **independent** if the events $\{X_1(\zeta) \leq x_1\}$ and $\{X_2(\zeta) \leq x_2\}$ are jointly independent; that is, the events do not influence one another, and

$$\Pr\left(X_1(\zeta) \leq x_1,\, X_2(\zeta) \leq x_2\right) = \Pr\left(X_1(\zeta) \leq x_1\right) \Pr\left(X_2(\zeta) \leq x_2\right)$$

Independence

Two random variables, $X_1(\zeta)$ and $X_2(\zeta)$ are **independent** if the events $\{X_1(\zeta) \leq x_1\}$ and $\{X_2(\zeta) \leq x_2\}$ are jointly independent; that is, the events do not influence one another, and

$$\Pr\left(X_1(\zeta) \leq x_1,\, X_2(\zeta) \leq x_2\right) = \Pr\left(X_1(\zeta) \leq x_1\right)\Pr\left(X_2(\zeta) \leq x_2\right)$$

This then implies that

$$F_{X_1,X_2}\left(x_1,\, x_2\right) = F_{X_1}\left(x_1\right) F_{X_2}\left(x_2\right)$$

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

Independence

Two random variables, $X_1(\zeta)$ and $X_2(\zeta)$ are **independent** if the events $\{X_1(\zeta) \leq x_1\}$ and $\{X_2(\zeta) \leq x_2\}$ are jointly independent; that is, the events do not influence one another, and

$$\Pr\left(X_1(\zeta) \leq x_1,\, X_2(\zeta) \leq x_2\right) = \Pr\left(X_1(\zeta) \leq x_1\right) \Pr\left(X_2(\zeta) \leq x_2\right)$$

This then implies that

$$F_{X_1,X_2}\left(x_1,\, x_2\right) = F_{X_1}\left(x_1\right) F_{X_2}\left(x_2\right)$$
$$f_{X_1,X_2}\left(x_1,\, x_2\right) = f_{X_1}\left(x_1\right) f_{X_2}\left(x_2\right)$$

- If the regions of support of the pdfs of $X(\zeta)$ and $Y(\zeta)$ are bounded, then $X(\zeta)$ and $Y(\zeta)$ cannot be independent if their ranges are dependent.

Independence

Example (Testing independence). Suppose the joint-pdf of two RVs $X(\zeta)$ and $Y(\zeta)$ is given by $f_{XY}(x,\, y) = 1 + xy$ for $0 < x < 1$ and $0 < y < 1$. Are $X(\zeta)$ and $Y(\zeta)$ independent?

Independence

Example (Testing independence). Suppose the joint-pdf of two RVs $X(\zeta)$ and $Y(\zeta)$ is given by $f_{XY}(x, y) = 1 + xy$ for $0 < x < 1$ and $0 < y < 1$. Are $X(\zeta)$ and $Y(\zeta)$ independent?

SOLUTION. The joint-pdf cannot be written in the form $g(x)\,h(x)$ for any functions g and h. Therefore, these RVs are not independent.

Independence

Example (Testing independence). Suppose the joint-pdf of two RVs $X(\zeta)$ and $Y(\zeta)$ is given by $f_{XY}(x, y) = 1 + xy$ for $0 < x < 1$ and $0 < y < 1$. Are $X(\zeta)$ and $Y(\zeta)$ independent?

SOLUTION. The joint-pdf cannot be written in the form $g(x)\,h(x)$ for any functions g and h. Therefore, these RVs are not independent.

Example (Testing independence). Let $f_{XY}(x, y) = 6x$ for $0 < x < y < 1$. Plot the region of support and determine if $X(\zeta)$ and $Y(\zeta)$ are independent.

Independence

- As an example that will be used many times in estimation theory, suppose that NRVs, $X_n(\zeta)$ for $n \in \{0, \dots, N-1\}$, are independent, and each have pdf given by $f_{X_n}(x_n)$.
- Then the joint-pdf of $\mathbf{X}(\zeta) = [X_0(\zeta), \cdots, X_N(\zeta)]^T$ is:

Independence

- As an example that will be used many times in estimation theory, suppose that N RVs, $X_n(\zeta)$ for $n \in \{0, \dots, N-1\}$, are independent, and each have pdf given by $f_{X_n}(x_n)$.
- Then the joint-pdf of $\mathbf{X}(\zeta) = [X_0(\zeta), \cdots, X_N(\zeta)]^T$ is:

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{n=0}^{N-1} f_{X_n}(x_n)$$

Independence

- As an example that will be used many times in estimation theory, suppose that NRVs, $X_n(\zeta)$ for $n \in \{0, \dots, N-1\}$, are independent, and each have pdf given by $f_{X_n}(x_n)$.
- Then the joint-pdf of $\mathbf{X}(\zeta) = [X_0(\zeta), \cdots, X_N(\zeta)]^T$ is:

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{n=0}^{N-1} f_{X_n}(x_n)$$

For example, suppose that $X_n(\zeta)$ is Gaussian distributed:

$$f_{X_n}(x_n) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x_n^2}{2}}$$

then:

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_n^2}{2}} = \frac{1}{(2\pi)^{\frac{N}{2}}} e^{-\frac{1}{2}\sum_{n=0}^{N-1} x_n^2}$$

Conditionals and Bayes's

The notion of joint probabilities and pdf also leads to the notion of conditional probabilities; what is the probability of a random vector $\mathbf{Y}(\zeta)$, given the random vector $\mathbf{X}(\zeta)$.

Conditionals and Bayes's

The notion of joint probabilities and pdf also leads to the notion of conditional probabilities; what is the probability of a random vector $\mathbf{Y}(\zeta)$, given the random vector $\mathbf{X}(\zeta)$.

The **conditional pdf** of $\mathbf{Y}(\zeta)$ given $\mathbf{X}(\zeta)$ is defined as:

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x}) = \frac{f_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y})}{f_{\mathbf{X}}(\mathbf{x})}$$

Conditionals and Bayes's

The notion of joint probabilities and pdf also leads to the notion of conditional probabilities; what is the probability of a random vector $\mathbf{Y}(\zeta)$, given the random vector $\mathbf{X}(\zeta)$.

The **conditional pdf** of $\mathbf{Y}(\zeta)$ given $\mathbf{X}(\zeta)$ is defined as:

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x}) = \frac{f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y})}{f_{\mathbf{X}}(\mathbf{x})}$$

If the random vectors $\mathbf{X}(\zeta)$ and $\mathbf{Y}(\zeta)$ are independent, then the conditional pdf must be identical to the unconditional pdf: $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x}) = f_{\mathbf{Y}}(\mathbf{y})$. Hence, it follows that:

$$f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) = f_{\mathbf{X}}(\mathbf{x})\, f_{\mathbf{Y}}(\mathbf{y})$$

Conditionals and Bayes's

Since

$$f_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x})\, f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x} \mid \mathbf{y})\, f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{Y}\mathbf{X}}(\mathbf{y}, \mathbf{x})$$

it follows

$$f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x} \mid \mathbf{y}) = \frac{f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x})\, f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{Y}}(\mathbf{y})}$$

Conditionals and Bayes's

Since

$$f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) = f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x})\, f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x} \mid \mathbf{y})\, f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{YX}}(\mathbf{y}, \mathbf{x})$$

it follows

$$f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x} \mid \mathbf{y}) = \frac{f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x})\, f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{Y}}(\mathbf{y})}$$

Since $f_{\mathbf{Y}}(\mathbf{y})$ can be expressed as:

$$f_{\mathbf{Y}}(\mathbf{y}) = \int_{\mathbb{R}} f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y})\, d\mathbf{x} = \int_{\mathbb{R}} f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x})\, f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}$$

then it follows

$$f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x} \mid \mathbf{y}) = \frac{f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x})\, f_{\mathbf{X}}(\mathbf{x})}{\int_{\mathbb{R}} f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x})\, f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}}$$

Conditionals and Bayes's

Example (Bayes's Theorem (Papoulis, Example 6-42)). An unknown random phase $\Theta(\zeta)$ is *a priori* assumed to be uniformly distributed in the interval $[0, 2\pi)$. The phase is observed through a noisy sensor, such that $R(\zeta) = \Theta(\zeta) + N(\zeta)$, where $N(\zeta)$ is Gaussian distributed with zero mean and variance σ_N^2.

What is the **posterior** pdf $f_{\Theta|R}(\theta \mid r)$?

Conditionals and Bayes's

Example (Bayes's Theorem (Papoulis, Example 6-42)). An unknown random phase $\Theta(\zeta)$ is *a priori* assumed to be uniformly distributed in the interval $[0, 2\pi)$. The phase is observed through a noisy sensor, such that $R(\zeta) = \Theta(\zeta) + N(\zeta)$, where $N(\zeta)$ is Gaussian distributed with zero mean and variance σ_N^2.

What is the **posterior** pdf $f_{\Theta|R}(\theta \mid r)$?

SOLUTION. In practical situations, it is reasonable to assume that $\Theta(\zeta)$ and $N(\zeta)$ are independent.

- Using the probability transformation rule, from $N(\zeta)$ to $R(\zeta) = \theta + N(\zeta)$ where $\Theta(\zeta) = \theta$ is considered fixed, it follows there is one inverse solution $n = r - \theta$, and the Jacobian of the transformation is unity. Therefore: □

Conditionals and Bayes's

Example (Bayes's Theorem (Papoulis, Example 6-42)). An unknown random phase $\Theta(\zeta)$ is *a priori* assumed to be uniformly distributed in the interval $[0, 2\pi)$. The phase is observed through a noisy sensor, such that $R(\zeta) = \Theta(\zeta) + N(\zeta)$, where $N(\zeta)$ is Gaussian distributed with zero mean and variance σ_N^2.

What is the **posterior** pdf $f_{\Theta|R}(\theta \mid r)$?

SOLUTION. In practical situations, it is reasonable to assume that $\Theta(\zeta)$ and $N(\zeta)$ are independent.

- Using the probability transformation rule, from $N(\zeta)$ to $R(\zeta) = \theta + N(\zeta)$ where $\Theta(\zeta) = \theta$ is considered fixed, it follows there is one inverse solution $n = r - \theta$, and the Jacobian of the transformation is unity. Therefore:

$$f_{R|\Theta}(r \mid \theta) = \frac{1}{1} f_N(r - \theta) = \frac{1}{\sqrt{2\pi\sigma_N^2}} e^{-\frac{(r-\theta)^2}{2\sigma_N^2}}$$

Conditionals and Bayes's

Example (Bayes's Theorem (Papoulis, Example 6-42)). SOLUTION. Using Bayes theorem, it directly follows that:

$$f_{\Theta|R}(\theta \mid r) = \frac{f_{R|\Theta}(r \mid \theta)\; f_\Theta(\theta)}{\int_0^{2\pi} f_{R|\Theta}\left(r \mid \hat{\theta}\right)\; f_\Theta\left(\hat{\theta}\right)\, d\hat{\theta}}$$

which, since $f_\Theta(\theta) = \frac{1}{2\pi}$ for $0 \le \theta < 2\pi$:

$$f_{\Theta|R}(\theta \mid r) = \frac{e^{-\frac{(r-\theta)^2}{2\sigma_N^2}}}{\int_0^{2\pi} e^{-\frac{(r-\theta)^2}{2\sigma_N^2}}\, d\theta} \quad 0 \le \theta < 2\pi \qquad \square$$

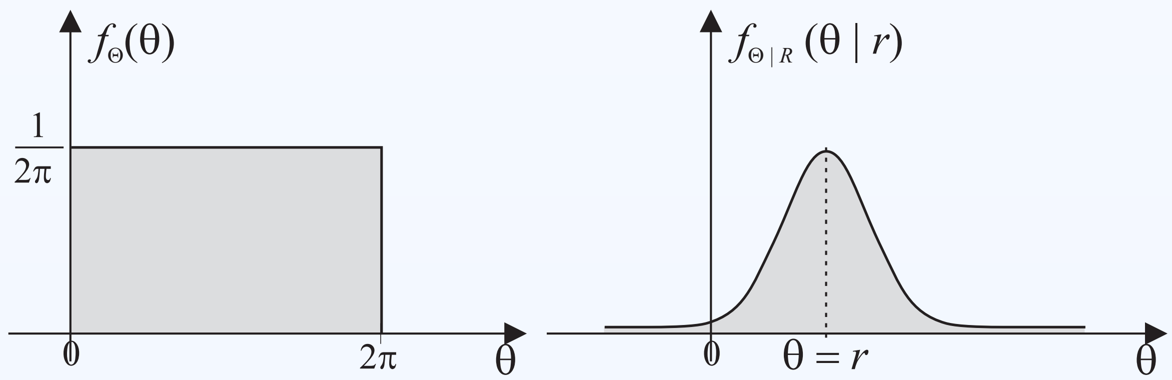
and zero otherwise, where it is noted that the factors $\frac{1}{2\pi}$ and $\frac{1}{\sqrt{2\pi\sigma_N^2}}$ have cancelled each other in the numerator and denominator.

Conditionals and Bayes's

Example (Bayes's Theorem (Papoulis, Example 6-42)). SOLUTION. Using Bayes theorem, it directly follows that:

$$f_{\Theta|R}(\theta \mid r) = \frac{e^{-\frac{(r-\theta)^2}{2\sigma_N^2}}}{\int_0^{2\pi} e^{-\frac{(r-\theta)^2}{2\sigma_N^2}}\, d\theta} \quad 0 \le \theta < 2\pi \qquad \square$$

Note the knowledge about the observation, r, is reflected in the posterior pdf of $\Theta(\zeta)$, and it shows higher probability density in the neighbourhood of $\Theta(\zeta) = r$.



Conditionals and Bayes's

Example (Chapman-Kolmogorov Equation). Consider a state-space model with an unknown state $\mathbf{x}_n$ and measurement vector $\mathbf{y}_n$.

Assume $p(\mathbf{x}_n \mid \mathbf{x}_{n-1}, \mathbf{y}_{1:n-1}) = p(\mathbf{x}_n \mid \mathbf{x}_{n-1})$ and $p(\mathbf{y}_n \mid \mathbf{x}_n, \mathbf{y}_{1:n-1}) = p(\mathbf{y}_n \mid \mathbf{x}_n)$.

Show that:

$$p(\mathbf{x}_n \mid \mathbf{y}_{1:n-1}) = \int p(\mathbf{x}_n \mid \mathbf{x}_{n-1})\, p(\mathbf{x}_{n-1} \mid \mathbf{y}_{1:n-1})\, d\mathbf{x}_{n-1}$$

$$p(\mathbf{x}_n \mid \mathbf{y}_{1:n}) = \frac{p(\mathbf{y}_n \mid \mathbf{x}_n)\, p(\mathbf{x}_n \mid \mathbf{y}_{1:n-1})}{p(\mathbf{y}_n \mid \mathbf{y}_{1:n-1})}$$

$\bowtie$

Conditionals and Bayes's

Example (Chapman-Kolmogorov Equation). Consider a state-space model with an unknown state $\mathbf{x}_n$ and measurement vector $\mathbf{y}_n$.

Assume $p\left(\mathbf{x}_n \mid \mathbf{x}_{n-1}, \mathbf{y}_{1:n-1}\right) = p\left(\mathbf{x}_n \mid \mathbf{x}_{n-1}\right)$ and $p\left(\mathbf{y}_n \mid \mathbf{x}_n, \mathbf{y}_{1:n-1}\right) = p\left(\mathbf{y}_n \mid \mathbf{x}_n\right)$.

SOLUTION. The first equation is a direct application of marginalisation of a joint-pdf:

$$\begin{aligned} p\left(\mathbf{x}_n \mid \mathbf{y}_{1:n-1}\right) &= \int p\left(\mathbf{x}_n, \mathbf{x}_{n-1} \mid \mathbf{y}_{1:n-1}\right) d\mathbf{x}_{n-1} \\ &= \int p\left(\mathbf{x}_n \mid \mathbf{x}_{n-1}, \mathbf{y}_{1:n-1}\right) p\left(\mathbf{x}_{n-1} \mid \mathbf{y}_{1:n-1}\right) d\mathbf{x}_{n-1} \\ &= \int p\left(\mathbf{x}_n \mid \mathbf{x}_{n-1}\right) p\left(\mathbf{x}_{n-1} \mid \mathbf{y}_{1:n-1}\right) d\mathbf{x}_{n-1} \qquad \square \end{aligned}$$

using the Markov property.

Conditionals and Bayes's

Example (Chapman-Kolmogorov Equation). Consider a state-space model with an unknown state $\mathbf{x}_n$ and measurement vector $\mathbf{y}_n$.

Assume $p(\mathbf{x}_n \mid \mathbf{x}_{n-1}, \mathbf{y}_{1:n-1}) = p(\mathbf{x}_n \mid \mathbf{x}_{n-1})$ and $p(\mathbf{y}_n \mid \mathbf{x}_n, \mathbf{y}_{1:n-1}) = p(\mathbf{y}_n \mid \mathbf{x}_n)$.

SOLUTION. The second equation is a direct application of Bayes's theorem keeping $\mathbf{y}_{1:n-1}$ a conditional in each term:

$$p(\mathbf{x}_n \mid \mathbf{y}_{1:n}) = p(\mathbf{x}_n \mid \mathbf{y}_n, \mathbf{y}_{1:n-1})$$

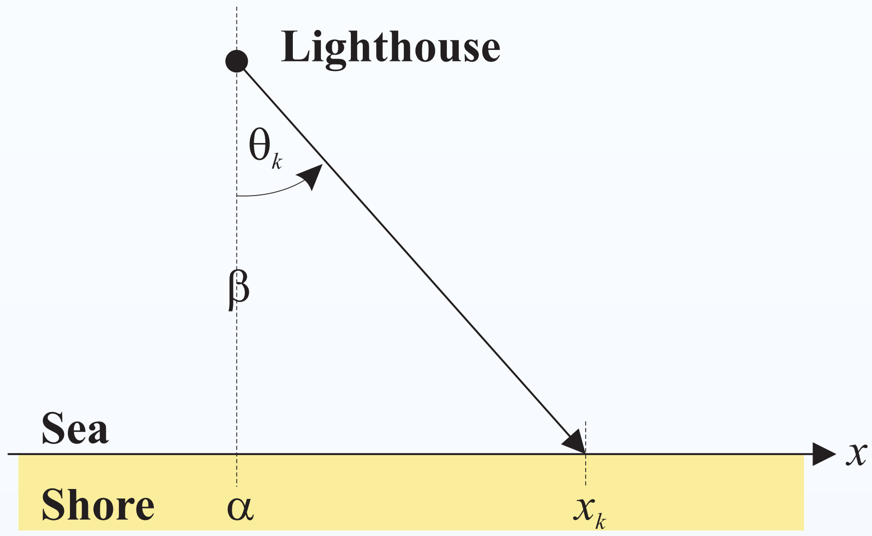
$$= \frac{p(\mathbf{y}_n \mid \mathbf{x}_n, \mathbf{y}_{1:n-1})\, p(\mathbf{x}_n \mid \mathbf{y}_{1:n-1})}{p(\mathbf{y}_n \mid \mathbf{y}_{1:n-1})} \qquad \square$$

Conditionals and Bayes's

– End-of-Topic 26: **Independence, Conditionals, and Bayes's Theorem Revisited** –

Any Questions?

Conditionals and Bayes's

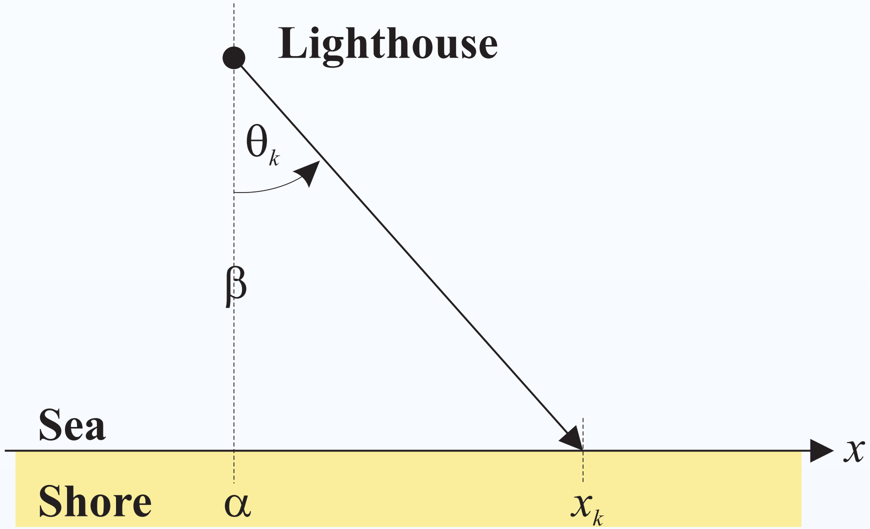


Example (Gull's lighthouse problem). A lighthouse is off a straight coastline at position α along the shore and distance β out at sea.

- It emits a series of short highly collimated flashes (i.e. a single ray of light) at random intervals and hence at random azimuths (i.e. the angle at which the light ray is emitted).
- These are intercepted on the coast by detectors that record that a flash occurred, but not the angle of arrival.
- N flashes recorded at $\{x_k\}$. Where is the lighthouse?

$\bowtie$

Conditionals and Bayes's



Example (Gull's lighthouse problem). This problem can be phrased in a number of other ways, such as throwing darts randomly at a wall and so forth. It is essentially a tomography problem, and is a classic inverse problem.

It can also be phrased as a geolocation problem, and there are a number of articles on this topic if you search the web!

Conditionals and Bayes's

Example (Gull's lighthouse problem). SOLUTION. Assign a uniform pdf to the azimuth of the observation which is given by θ. Hence,

$$f_\Theta(\theta) = \begin{cases} \frac{1}{\pi} & -\frac{\pi}{2} < \theta < \frac{\pi}{2} \\ 0 & \text{otherwise} \end{cases}$$

□

Conditionals and Bayes's

Example (Gull's lighthouse problem). SOLUTION. Assign a uniform pdf to the azimuth of the observation which is given by θ. Hence,

$$f_\Theta(\theta) = \begin{cases} \frac{1}{\pi} & -\frac{\pi}{2} < \theta < \frac{\pi}{2} \\ 0 & \text{otherwise} \end{cases}$$

- Since the photo-detectors are only sensitive to position along the coast rather than direction, it is necessary to relate θ to x:

$$\beta \tan\theta = x - \alpha$$

□

Conditionals and Bayes's

Example (Gull's lighthouse problem). SOLUTION. Assign a uniform pdf to the azimuth of the observation which is given by θ. Hence,

$$f_\Theta(\theta) = \begin{cases} \frac{1}{\pi} & -\frac{\pi}{2} < \theta < \frac{\pi}{2} \\ 0 & \text{otherwise} \end{cases}$$

- Since the photo-detectors are only sensitive to position along the coast rather than direction, it is necessary to relate θ to x:

$$\beta \tan\theta = x - \alpha$$

- Using the probability transformation rule:

$$f_X(x \mid \alpha) = \frac{\beta}{\pi\left[\beta^2 + (x-\alpha)^2\right]}$$

□

Conditionals and Bayes's

Example (Gull's lighthouse problem). SOLUTION. Assuming observations are independent, the joint-pdf of all the data is:

$$f_{\mathbf{X}}\left(\mathbf{x} \,|\, \alpha\right) = f_{\mathbf{X}}\left(x_1, \dots, x_N \,|\, \alpha\right) = \prod_{k=1}^{N} f_X\left(x_k \,|\, \alpha\right)$$

$$= \prod_{k=1}^{N} \frac{\beta}{\pi\left[\beta^2 + (x_k - \alpha)^2\right]}$$

□

Conditionals and Bayes's

Example (Gull's lighthouse problem). SOLUTION. Assuming observations are independent, the joint-pdf of all the data is:

$$f_{\mathbf{X}}\left(\mathbf{x} \,|\, \alpha\right) = f_{\mathbf{X}}\left(x_1, \dots, x_N \,|\, \alpha\right) = \prod_{k=1}^{N} f_X\left(x_k \,|\, \alpha\right)$$

$$= \prod_{k=1}^{N} \frac{\beta}{\pi\left[\beta^2 + (x_k - \alpha)^2\right]}$$

The position of the lighthouse is then expressed by:

$$f_A\left(\alpha \,|\, \mathbf{x}\right) = \frac{f_{\mathbf{X}}\left(\mathbf{x} \,|\, \alpha\right)\, f_A\left(\alpha\right)}{f_{\mathbf{X}}\left(\mathbf{x}\right)}$$

□

Conditionals and Bayes's

Example (Gull's lighthouse problem). SOLUTION. Assign a uniform for the *prior* for distance along the shore:

$$f_A(\alpha) = \begin{cases} \frac{1}{\alpha_{\max} - \alpha_{\min}} & \alpha_{\min} \le \alpha \le \alpha_{\max} \\ 0 & \text{otherwise} \end{cases}$$

$\square$

Conditionals and Bayes's

Example (Gull's lighthouse problem). SOLUTION. Assign a uniform for the *prior* for distance along the shore:

$$f_A(\alpha) = \begin{cases} \frac{1}{\alpha_{\max}-\alpha_{\min}} & \alpha_{\min} \le \alpha \le \alpha_{\max} \\ 0 & \text{otherwise} \end{cases}$$

Hence:

$$f_A(\alpha \mid \mathbf{x}) = \frac{f_{\mathbf{X}}(\mathbf{x} \mid \alpha)\, f_A(\alpha)}{f_{\mathbf{X}}(\mathbf{x})} \propto f_{\mathbf{X}}(\mathbf{x} \mid \alpha)\, f_A(\alpha)$$

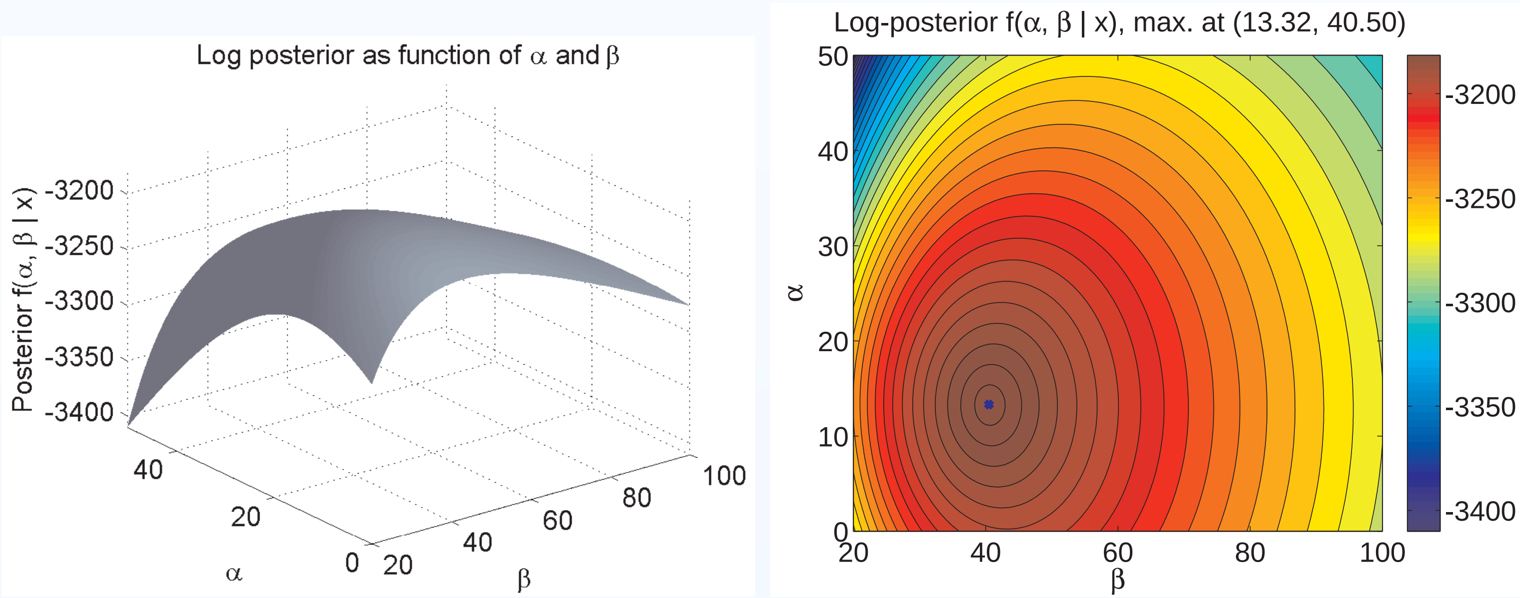
$$\propto \frac{1}{\alpha_{\max} - \alpha_{\min}} \prod_{k=1}^{N} \frac{\beta}{\pi\left[\beta^2 + (x_k - \alpha)^2\right]}, \quad \text{for } \alpha_{\min} \le \alpha \le \alpha_{\max}$$

and zero otherwise. Hence, this **posterior density** can be maximised to find the best estimate of α.

Conditionals and Bayes's

Example (Gull's lighthouse problem). SOLUTION.

$$f_A(\alpha) = \begin{cases} \frac{1}{\alpha_{\max}-\alpha_{\min}} & \alpha_{\min} \le \alpha \le \alpha_{\max} \\ 0 & \text{otherwise} \end{cases} \qquad \square$$



Exhaustive Evaluation of Log-posterior.

Conditionals and Bayes's

This example highlights two key problems in Signal Processing:

Integration Marginalising out nuisance parameters:

$$f_A(\alpha \,|\, \mathbf{x}) = \int f_A(\alpha, \beta \,|\, \mathbf{x}) \, d\beta$$

Optimisation Finding the maximum marginal *a posteriori* (MMAP) estimate:

$$\hat{\alpha} = \arg_\alpha \max f_A(\alpha \,|\, \mathbf{x})$$

Conditionals and Bayes's

– End-of-Topic 27: **Tomography: An Inverse Problem using Probability Transformations, Conditional Probability, Independence, Bayes Theorem, Marginalisation, and Optimisation.** –

Any Questions?

Probability Transformation Rule

Theorem (Probability Transformation Rule). The set of random variables $\mathbf{X}(\zeta) = \{X_n(\zeta),\, n \in \mathcal{N}\}$ are transformed to a new set of RVs, $\mathbf{Y}(\zeta) = \{Y_n(\zeta),\, n \in \mathcal{N}\}$, using the transformations:

$$Y_n(\zeta) = g_n(\mathbf{X}(\zeta)), \quad n \in \mathcal{N} \qquad \diamondsuit$$

where $\mathbf{g}(\cdot)$ denotes a vector of functions $Y_n(\zeta) = g_n(\mathbf{X}(\zeta))$.

Probability Transformation Rule

Theorem (Probability Transformation Rule). The set of random variables $\mathbf{X}(\zeta) = \{X_n(\zeta),\, n \in \mathcal{N}\}$ are transformed to a new set of RVs, $\mathbf{Y}(\zeta) = \{Y_n(\zeta),\, n \in \mathcal{N}\}$, using the transformations:

$$Y_n(\zeta) = g_n(\mathbf{X}(\zeta)), \quad n \in \mathcal{N}$$

where $\mathbf{g}(\cdot)$ denotes a vector of functions $Y_n(\zeta) = g_n(\mathbf{X}(\zeta))$.

Assuming M-real vector-roots of the equation $\mathbf{y} = \mathbf{g}(\mathbf{x})$ by $\{\mathbf{x}_m,\, m \in \mathcal{M}\}$,

$$\mathbf{y} = \mathbf{g}(\mathbf{x}_1) = \cdots = \mathbf{g}(\mathbf{x}_M)$$

then the joint-pdf of $\mathbf{Y}(\zeta)$ in terms of (i. t. o.) the joint-pdf of $\mathbf{X}(\zeta)$ is:

$$f_{\mathbf{Y}}(\mathbf{y}) = \sum_{m=1}^{M} \frac{f_{\mathbf{X}}(\mathbf{x}_m)}{|J(\mathbf{x}_m)|}$$

Probability Transformation Rule

Theorem (Probability Transformation Rule). The **Jacobian** of the transformation, $J_{\mathbf{g}}(\mathbf{x})$, is given by:

$$J_{\mathbf{g}}(\mathbf{x}) \triangleq \frac{\partial(y_1, \dots, y_N)}{\partial(x_1, \dots, x_N)} = \begin{vmatrix} \frac{\partial g_1(\mathbf{x})}{\partial x_1} & \frac{\partial g_2(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial g_N(\mathbf{x})}{\partial x_1} \\ \frac{\partial g_1(\mathbf{x})}{\partial x_2} & \frac{\partial g_2(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial g_N(\mathbf{x})}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_1(\mathbf{x})}{\partial x_N} & \frac{\partial g_2(\mathbf{x})}{\partial x_N} & \cdots & \frac{\partial g_N(\mathbf{x})}{\partial x_N} \end{vmatrix}$$

♢

Probability Transformation Rule

Theorem (Probability Transformation Rule). The **Jacobian** of the transformation, $J_{\mathbf{g}}(\mathbf{x})$, is given by:

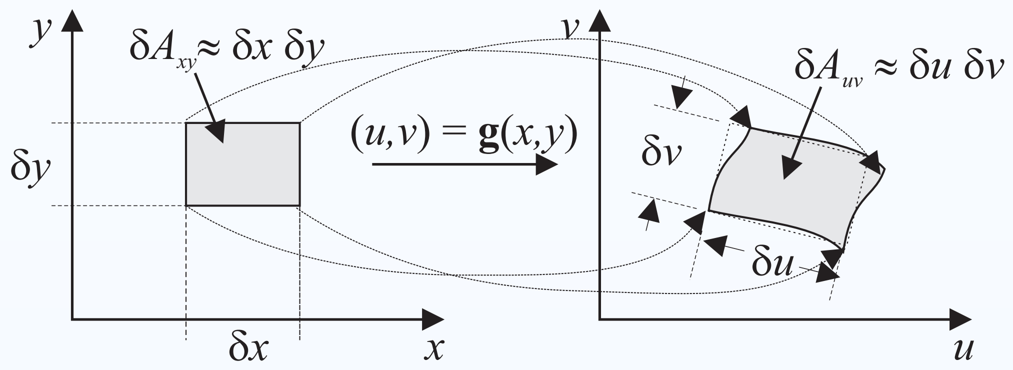
$$J_{\mathbf{g}}(\mathbf{x}) \triangleq \frac{\partial(y_1, \dots, y_N)}{\partial(x_1, \dots, x_N)} = \begin{vmatrix} \frac{\partial g_1(\mathbf{x})}{\partial x_1} & \frac{\partial g_2(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial g_N(\mathbf{x})}{\partial x_1} \\ \frac{\partial g_1(\mathbf{x})}{\partial x_2} & \frac{\partial g_2(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial g_N(\mathbf{x})}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_1(\mathbf{x})}{\partial x_N} & \frac{\partial g_2(\mathbf{x})}{\partial x_N} & \cdots & \frac{\partial g_N(\mathbf{x})}{\partial x_N} \end{vmatrix}$$

From vector calculus, the Jacobian can also be expressed as:

$$\frac{1}{J_{\mathbf{g}}(\mathbf{x})} \triangleq \frac{\partial(x_1, \dots, x_N)}{\partial(y_1, \dots, y_N)} = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_2}{\partial y_1} & \cdots & \frac{\partial x_N}{\partial y_1} \\ \frac{\partial x_1}{\partial y_2} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_N}{\partial y_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_1}{\partial y_N} & \frac{\partial x_2}{\partial y_N} & \cdots & \frac{\partial x_N}{\partial y_N} \end{vmatrix} \qquad \diamondsuit$$

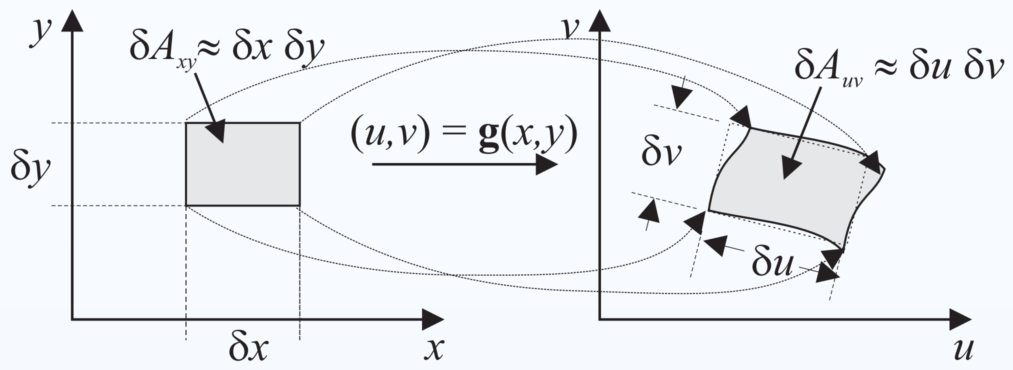
Probability Transformation Rule

The Jacobian determinant represents how an elemental region in one domain changes volume when mapped to another domain.



Probability Transformation Rule

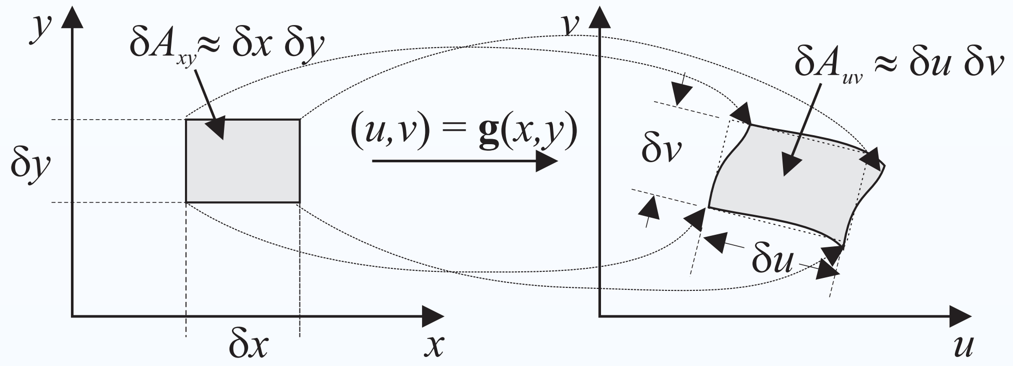
The Jacobian determinant represents how an elemental region in one domain changes volume when mapped to another domain.



- This elemental area is mapped into the (u, v) domain through the relationships $u = g_1(x, y)$ and $v = g_2(x, y)$.
- The Jacobian indicates the ratio of these two areas:

$$\delta A_{uv} \approx J_{xy\to uv}\, \delta A_{xy} \qquad J_{xy\to uv} \approx \frac{\delta u\, \delta v}{\delta x\, \delta y}$$

Probability Transformation Rule



- This elemental area is mapped into the (u, v) domain through the relationships $u = g_1(x, y)$ and $v = g_2(x, y)$.
- The Jacobian indicates the ratio of these two areas:

$$\delta A_{uv} \approx J_{xy\to uv}\, \delta A_{xy} \qquad J_{xy\to uv} \approx \frac{\delta u\, \delta v}{\delta x\, \delta y}$$

In the limit, it can be shown that the Jacobian determinant is:

$$J_{uv\to xy} = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial v}{\partial x} \\ \frac{\partial u}{\partial y} & \frac{\partial v}{\partial y} \end{vmatrix}$$

Polar Transformation

Consider the transformation from the random vector $\mathbf{C}(\zeta) = [X(\zeta)\,,\, Y(\zeta)]^T$ to $\mathbf{P}(\zeta) = [r(\zeta),\, \theta(\zeta)]^T$, where

$$r(\zeta) = \sqrt{X^2(\zeta) + Y^2(\zeta)}$$

$$\theta(\zeta) = \arctan \frac{Y(\zeta)}{X(\zeta)}$$

Polar Transformation

Consider the transformation from the random vector $\mathbf{C}(\zeta) = [X(\zeta)\,,\,Y(\zeta)]^T$ to $\mathbf{P}(\zeta) = [r(\zeta),\,\theta(\zeta)]^T$, where

$$r(\zeta) = \sqrt{X^2(\zeta) + Y^2(\zeta)}$$

$$\theta(\zeta) = \arctan \frac{Y(\zeta)}{X(\zeta)}$$

The Jacobian is given by:

$$J_{\mathbf{g}}(\mathbf{c}) = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix}^{-1} = \frac{1}{r}$$

Thus, it follows that:

$$f_{R,\Theta}\left(r,\theta\right) = r\, f_{XY}\left(r\,\cos\theta,\, r\,\sin\theta\right)$$

Polar Transformation

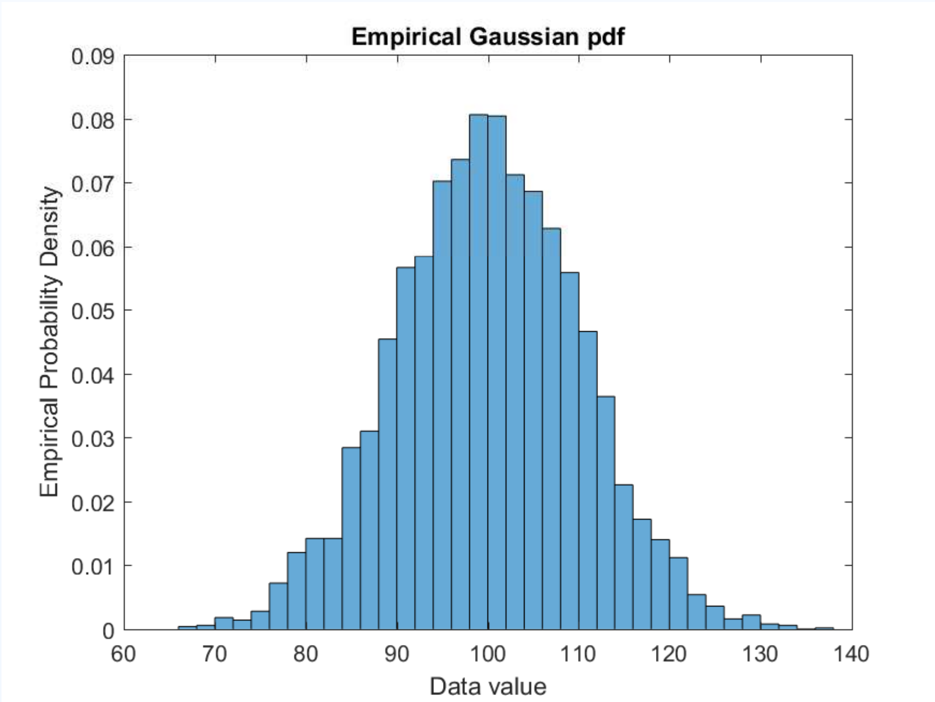
– End-of-Topic 28: **Probability Transformation rule for Random Vectors** –

Any Questions?

Generating Gaussian distributed samples

It is often important to generate samples from a Gaussian density, primarily for simulation studies.

- In practice, it is difficult for a computer to generate random numbers from an arbitrary density.
- However, it is possible to generate uniform random variates fairly easily.



Generating Gaussian distributed samples

Consider the transformation between two uniform random variables

$$f_{X_k}(x_k) = \mathbb{I}_{0,1}(x_k), \quad k = 1, 2$$

where $\mathbb{I}_{\mathcal{A}}(x) = 1$ if $x \in \mathcal{A}$, and zero otherwise.

Generating Gaussian distributed samples

Consider the transformation between two uniform random variables

$$f_{X_k}(x_k) = \mathbb{I}_{0,1}(x_k), \quad k = 1, 2$$

where $\mathbb{I}_{\mathcal{A}}(x) = 1$ if $x \in \mathcal{A}$, and zero otherwise.

Now let two random variables y_1, y_2 be given by:

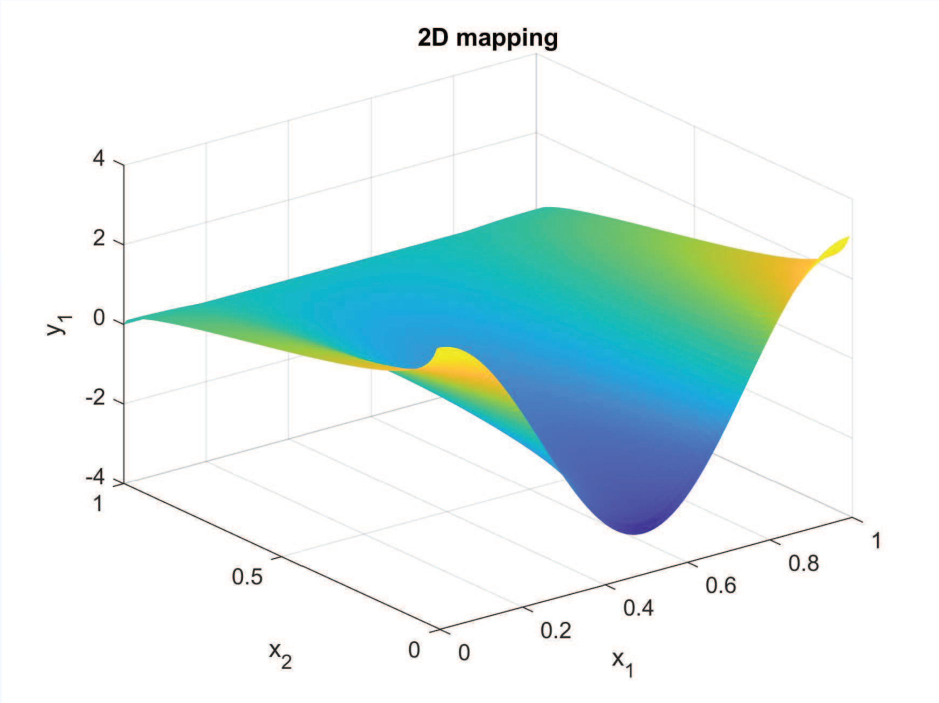
$$y_1 = \sqrt{-2 \ln x_1} \cos 2\pi x_2$$

$$y_2 = \sqrt{-2 \ln x_1} \sin 2\pi x_2$$

Generating Gaussian distributed samples

Now let two random variables y_1, y_2 be given by:

$$y_1 = \sqrt{-2\ln x_1}\,\cos 2\pi x_2$$
$$y_2 = \sqrt{-2\ln x_1}\,\sin 2\pi x_2$$



Generating Gaussian distributed samples

It follows, by rearranging these equations, that:

$$x_1 = \exp\left[-\frac{1}{2}(y_1^2 + y_2^2)\right]$$

$$x_2 = \frac{1}{2\pi}\arctan\frac{y_2}{y_1}$$

Generating Gaussian distributed samples

It follows, by rearranging these equations, that:

$$x_1 = \exp\left[-\frac{1}{2}(y_1^2 + y_2^2)\right]$$

$$x_2 = \frac{1}{2\pi} \arctan \frac{y_2}{y_1}$$

The Jacobian determinant can be calculated as:

$$J(x_1, x_2) = \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{vmatrix}$$

$$= \begin{vmatrix} \frac{-1}{x_1\sqrt{-2\ln x_1}} \cos 2\pi x_2 & -2\pi\sqrt{-2\ln x_1} \sin 2\pi x_2 \\ \frac{-1}{x_1\sqrt{-2\ln x_1}} \sin 2\pi x_2 & 2\pi\sqrt{-2\ln x_1} \cos 2\pi x_2 \end{vmatrix}$$

$$= \frac{2\pi}{x_1}$$

Generating Gaussian distributed samples

Hence, it follows:

$$f_Y(y_1, y_2) = \frac{x_1}{2\pi} = \left[\frac{1}{\sqrt{2\pi}} e^{-y_1^2/2}\right] \left[\frac{1}{\sqrt{2\pi}} e^{-y_2^2/2}\right]$$

- since the domain $[0, 1]^2$ is mapped to the range $(-\infty, \infty)^2$, thus covering the range of real numbers.
- This is the product of the pdfs of y_1 and y_2, and therefore each y_k is independent and identically distributed (i. i. d.) according to the normal distribution

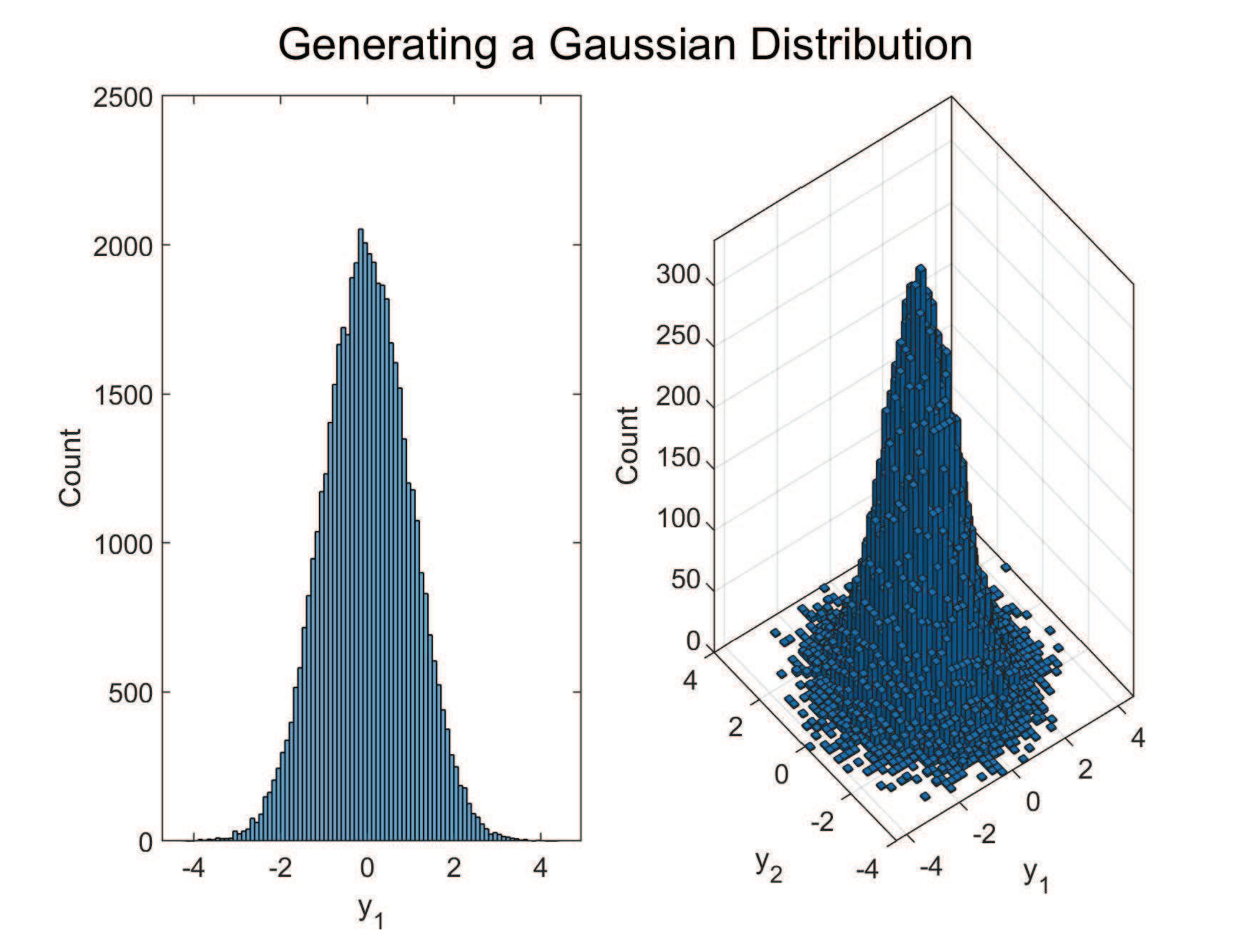
Generating Gaussian distributed samples

Hence, it follows:

$$f_Y(y_1, y_2) = \frac{x_1}{2\pi} = \left[\frac{1}{\sqrt{2\pi}} e^{-y_1^2/2}\right]\left[\frac{1}{\sqrt{2\pi}} e^{-y_2^2/2}\right]$$

- since the domain $[0, 1]^2$ is mapped to the range $(-\infty, \infty)^2$, thus covering the range of real numbers.
- This is the product of the pdfs of y_1 and y_2, and therefore each y_k is i. i. d. according to the normal distribution

Generating Gaussian distributed samples



The resulting histogram from the generation of these Gaussian samples.

Generating Gaussian distributed samples

– End-of-Topic 29: **Generating Gaussian Samples** –

Any Questions?

Auxiliary Variables

- So far transforming from N RVs to N RVs considered.
- However, what about the case of transforming from N RVs to M RVs, where $M < N$; for example, $Z(\zeta) = g(X(\zeta), Y(\zeta))$?

Auxiliary Variables

- So far transforming from N RVs to N RVs considered.
- However, what about the case of transforming from N RVs to M RVs, where $M < N$; for example, $Z(\zeta) = g(X(\zeta), Y(\zeta))$?

The density of a RV that is *one* function $Z(\zeta) = g(X(\zeta), Y(\zeta))$ of two RVs can be determined by choosing an **auxiliary variable**, $W(\zeta)$. Examples might be $W(\zeta) = X(\zeta)$ or $W(\zeta) = Y(\zeta)$.

Auxiliary Variables

- So far transforming from NRVs to NRVs considered.
- However, what about the case of transforming from NRVs to MRVs, where $M < N$; for example, $Z(\zeta) = g(X(\zeta), Y(\zeta))$?

The density of a RV that is *one* function $Z(\zeta) = g(X(\zeta), Y(\zeta))$ of two RVs can be determined by choosing an **auxiliary variable**, $W(\zeta)$. Examples might be $W(\zeta) = X(\zeta)$ or $W(\zeta) = Y(\zeta)$.

The density of $Z(\zeta)$ is found by the **probability transformation**,

$$f_{WZ}(w, z)\, dw = \sum_{m=1}^{M} \frac{f_{\mathbf{XY}}(x_m, y_m)}{|J(x_m, y_m)|}$$

Auxiliary Variables

- So far transforming from NRVs to NRVs considered.
- However, what about the case of transforming from NRVs to MRVs, where $M < N$; for example, $Z(\zeta) = g(X(\zeta), Y(\zeta))$?

The density of a RV that is *one* function $Z(\zeta) = g(X(\zeta), Y(\zeta))$ of two RVs can be determined by choosing an **auxiliary variable**, $W(\zeta)$. Examples might be $W(\zeta) = X(\zeta)$ or $W(\zeta) = Y(\zeta)$.

The density of $Z(\zeta)$ is found by the **probability transformation**,

$$f_{WZ}(w, z)\, dw = \sum_{m=1}^{M} \frac{f_{\mathbf{XY}}(x_m, y_m)}{|J(x_m, y_m)|}$$

followed by **marginalisation**:

$$f_Z(z) = \int_{\mathbb{R}} f_{WZ}(w, z)\, dw = \sum_{m=1}^{M} \int_{\mathbb{R}} \frac{f_{\mathbf{XY}}(x_m, y_m)}{|J(x_m, y_m)|}\, dw$$

Auxiliary Variables

Example (Sum of two RVs). If $X(\zeta)$ and $Y(\zeta)$ have joint-pdf $f_{XY}(x, y)$, find the pdf of the RV $Z(\zeta) = aX(\zeta) + bY(\zeta)$.

Auxiliary Variables

Example (Sum of two RVs). If $X(\zeta)$ and $Y(\zeta)$ have joint-pdf $f_{XY}(x, y)$, find the pdf of the RV $Z(\zeta) = aX(\zeta) + bY(\zeta)$.

SOLUTION. Use as the auxiliary variable the function $W(\zeta) = Y(\zeta)$. The system $z = ax + by$, $w = y$ has a single solution at $x = \frac{z-bw}{a}$, $y = w$.

Auxiliary Variables

Example (Sum of two RVs). If $X(\zeta)$ and $Y(\zeta)$ have joint-pdf $f_{XY}(x, y)$, find the pdf of the RV $Z(\zeta) = aX(\zeta) + bY(\zeta)$.

SOLUTION. Use as the auxiliary variable the function $W(\zeta) = Y(\zeta)$. The system $z = ax + by$, $w = y$ has a single solution at $x = \frac{z-bw}{a}$, $y = w$.

Hence, the Jacobian is given by:

$$J(x, y) = \begin{vmatrix} \frac{\partial w}{\partial x} & \frac{\partial z}{\partial x} \\ \frac{\partial w}{\partial y} & \frac{\partial z}{\partial y} \end{vmatrix} = \begin{vmatrix} 0 & a \\ 1 & b \end{vmatrix} = -a$$

□

Auxiliary Variables

Example (Sum of two RVs). If $X(\zeta)$ and $Y(\zeta)$ have joint-pdf $f_{XY}(x, y)$, find the pdf of the RV $Z(\zeta) = aX(\zeta) + bY(\zeta)$.

SOLUTION. Use as the auxiliary variable the function $W(\zeta) = Y(\zeta)$. The system $z = ax + by$, $w = y$ has a single solution at $x = \frac{z-bw}{a}$, $y = w$.

Hence, the Jacobian is given by:

$$J(x, y) = \begin{vmatrix} \frac{\partial w}{\partial x} & \frac{\partial z}{\partial x} \\ \frac{\partial w}{\partial y} & \frac{\partial z}{\partial y} \end{vmatrix} = \begin{vmatrix} 0 & a \\ 1 & b \end{vmatrix} = -a$$

Thus:

$$f_Z(z) = \frac{1}{|a|} \int_{\mathbb{R}} f_{XY}\left(\frac{z-bw}{a}, w\right) dw \qquad \square$$

Auxiliary Variables

Note that you might be concerned about the choice of the auxiliary variable, and what happens if you chose something different to that used here.

Auxiliary Variables

Note that you might be concerned about the choice of the auxiliary variable, and what happens if you chose something different to that used here.

- The answer is that, as long as the auxliary variable is a function of at least one of the RVs, then it doesn't really matter, as the **marginalisation** stage will usually yield the same answer.
- Nevertheless, it usally pays to chose the auxiliary variable carefully to minimise any difficulties in evaluating the marginal-pdf.

Auxiliary Variables

Note that you might be concerned about the choice of the auxiliary variable, and what happens if you chose something different to that used here.

- The answer is that, as long as the auxliary variable is a function of at least one of the RVs, then it doesn't really matter, as the **marginalisation** stage will usually yield the same answer.
- Nevertheless, it usally pays to chose the auxiliary variable carefully to minimise any difficulties in evaluating the marginal-pdf.

As an example, consider using $W(\zeta) = X(\zeta) - Y(\zeta)$ in the previous example).

Auxiliary Variables

Example ([Papoulis:1991, Page 149, Problem 6-8]). The RVs $X(\zeta)$ and $Y(\zeta)$ are independent with Rayleigh densities:

$$f_X(x) = \frac{x}{\alpha^2} \exp\left\{-\frac{x^2}{2\alpha^2}\right\} \mathbb{I}_{\mathbb{R}^+}(x)$$

$$f_Y(y) = \frac{y}{\beta^2} \exp\left\{-\frac{y^2}{2\beta^2}\right\} \mathbb{I}_{\mathbb{R}^+}(y)$$

1. Show that if $Z(\zeta) = {}^{X(\zeta)}/_{Y(\zeta)}$, then:

$$f_Z(z) = \frac{2\alpha^2}{\beta^2} \frac{z}{\left(z^2 + \frac{\alpha^2}{\beta^2}\right)^2} \mathbb{I}_{\mathbb{R}^+}(z)$$

2. Using this result, show that for any $k > 0$,

$$\Pr\left(X(\zeta) \le k\, Y(\zeta)\right) = \frac{k^2}{k^2 + \frac{\alpha^2}{\beta^2}}$$

⋈

Auxiliary Variables

– End-of-Topic 30: **Using auxiliary variables and their applications** –

Any Questions?

Statistical Description

Statistical averages are more manageable, but less of a complete description of random vectors.

Statistical Description

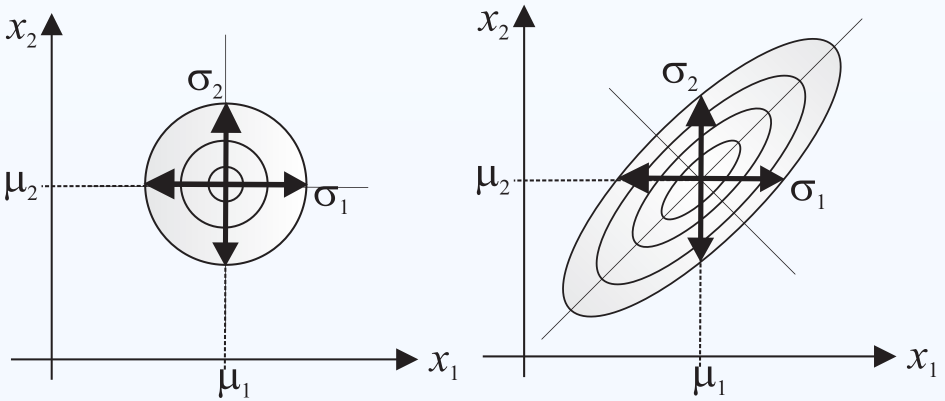
Statistical averages are more manageable, but less of a complete description of random vectors.

- With care, it is possible to extend many of the statistical descriptors for scalar RVs to random vectors.

Statistical Description

Statistical averages are more manageable, but less of a complete description of random vectors.

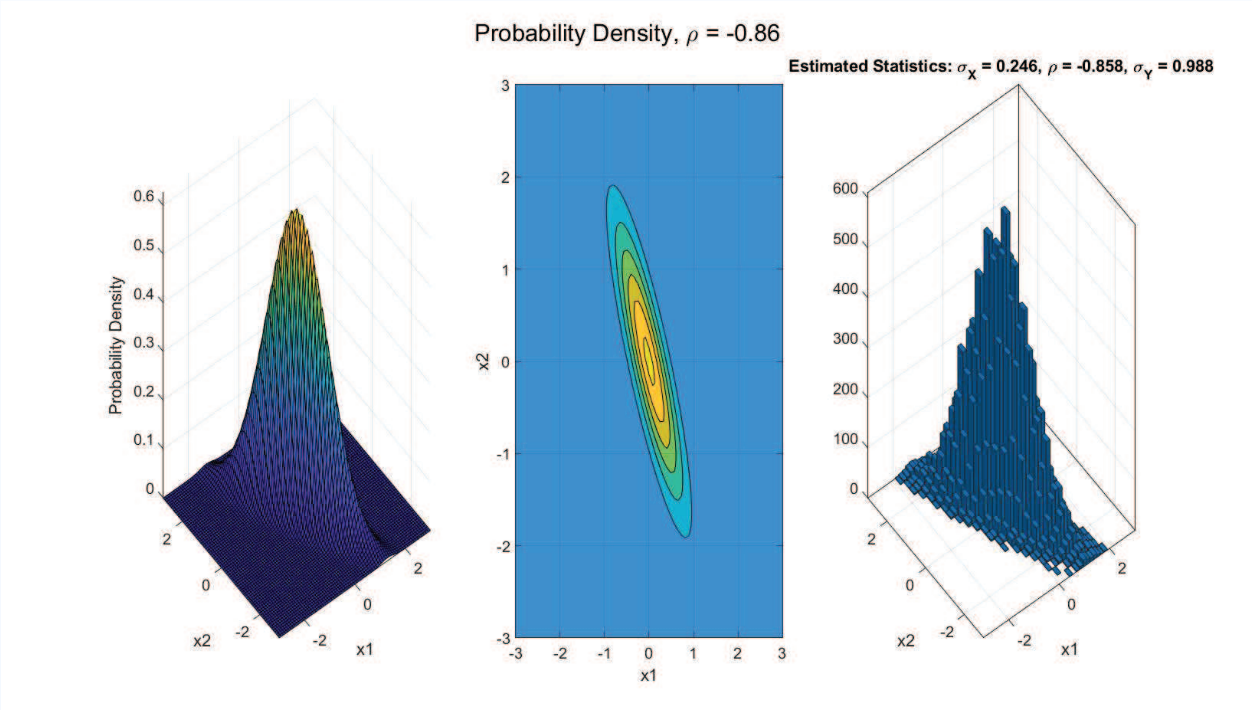
- With care, it is possible to extend many of the statistical descriptors for scalar RVs to random vectors.
- Second-order moments of individual RVs do not adequately capture key characteristics of the joint-pdf.



Mean and second-moments of individual RVs does not capture all of the information about the joint-pdf.

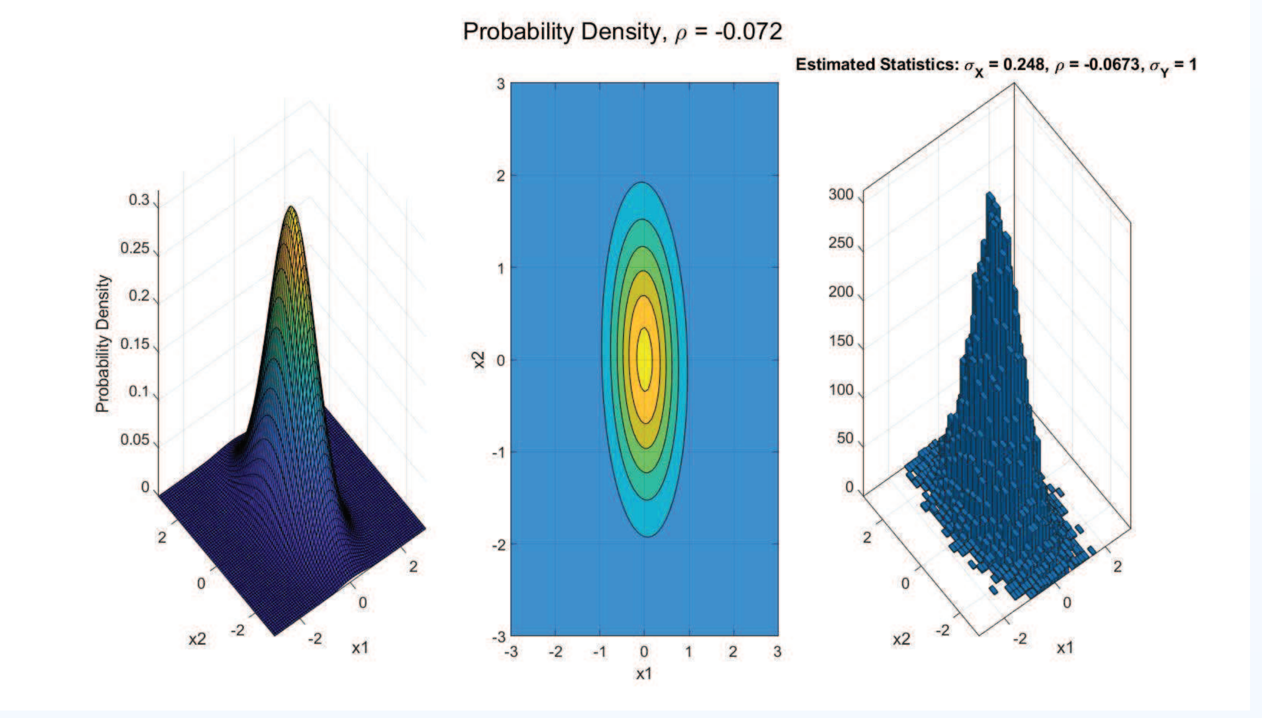
Statistical Description

Statistical averages are more manageable, but less of a complete description of random vectors.



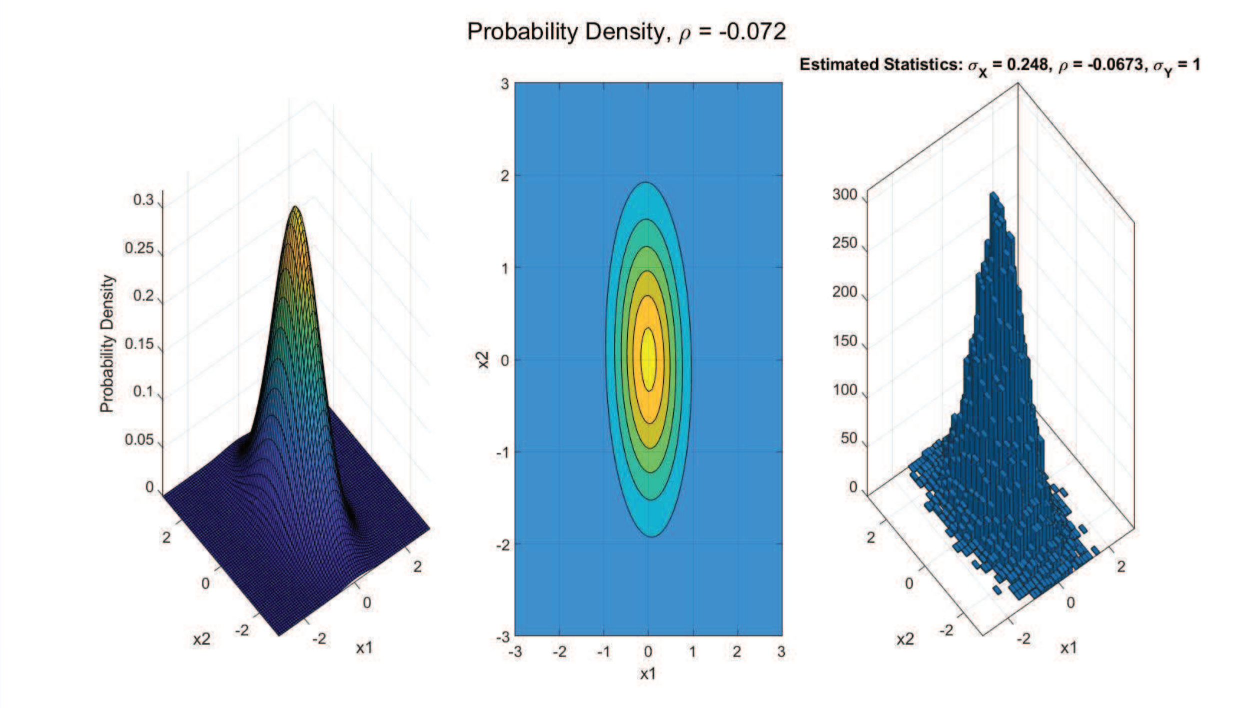
Statistical Description

Statistical averages are more manageable, but less of a complete description of random vectors.



Statistical Description

Statistical averages are more manageable, but less of a complete description of random vectors.



- Consequently, it is important to understand that multiple RVs leads to the notion of measuring their dependence. This concept is useful in abstract, but also for stochastic processes.

Mean Vectors and Correlation Matrices

Mean vector The **mean vector** is the first-moment of the random vector, and is given by:

$$\boldsymbol{\mu}_{\mathbf{X}} = \mathbb{E}\left[\mathbf{X}(\zeta)\right] = \begin{bmatrix} \mathbb{E}\left[X_1(\zeta)\right] \\ \vdots \\ \mathbb{E}\left[X_N(\zeta)\right] \end{bmatrix} = \begin{bmatrix} \mu_{X_1} \\ \vdots \\ \mu_{X_N} \end{bmatrix}$$

Mean Vectors and Correlation Matrices

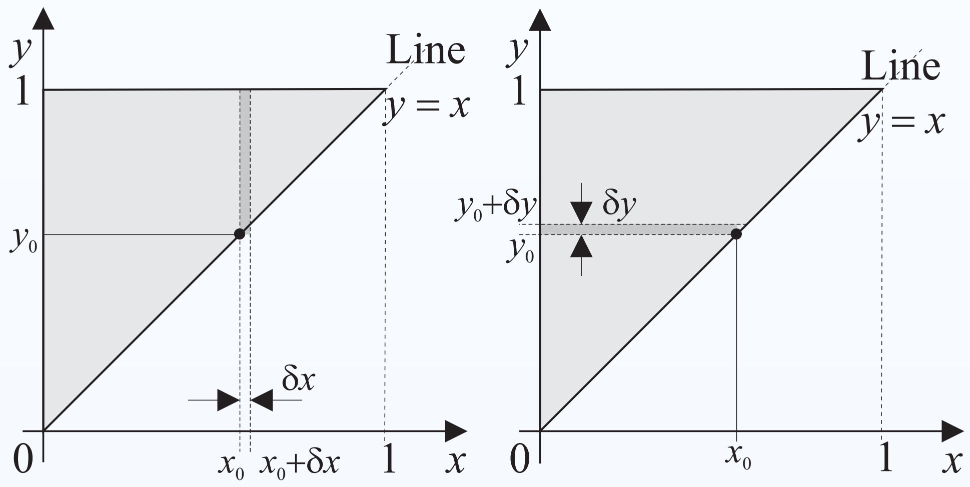
Mean vector The **mean vector** is the first-moment of the random vector, and is given by:

$$\boldsymbol{\mu}_{\mathbf{X}} = \mathbb{E}\left[\mathbf{X}(\zeta)\right] = \begin{bmatrix} \mathbb{E}\left[X_1(\zeta)\right] \\ \vdots \\ \mathbb{E}\left[X_N(\zeta)\right] \end{bmatrix} = \begin{bmatrix} \mu_{X_1} \\ \vdots \\ \mu_{X_N} \end{bmatrix}$$

Example (Mean Vector). Let $f_{XY}(x, y) = 2$ for $0 < x < y < 1$ and zero otherwise. Find the mean-vector.

Mean Vectors and Correlation Matrices

Mean vector The **mean vector** is the first-moment :

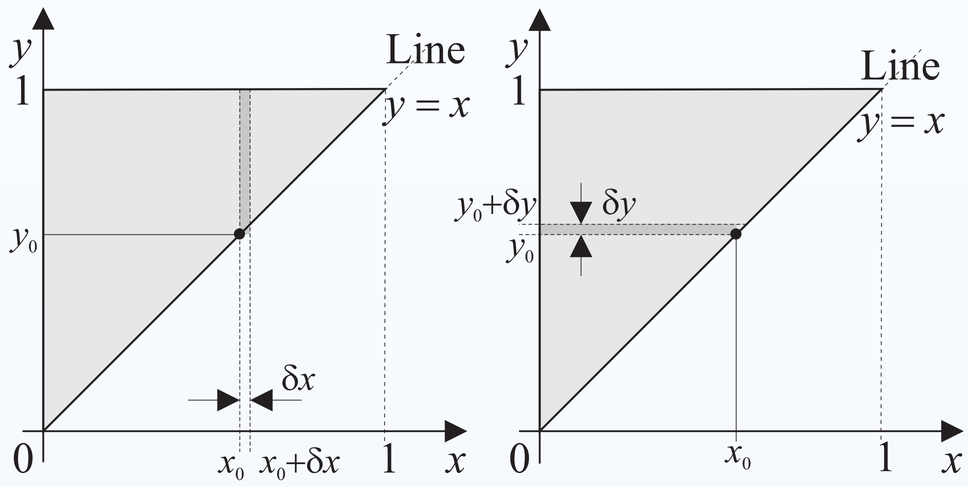


Example (Mean Vector). Let $f_{XY}(x, y) = 2$ for $0 < x < y < 1$ and zero otherwise. Find the mean-vector.

SOLUTION. The calculation involves finding the marginals and then the expected value.

Mean Vectors and Correlation Matrices

Mean vector The **mean vector** is the first-moment :



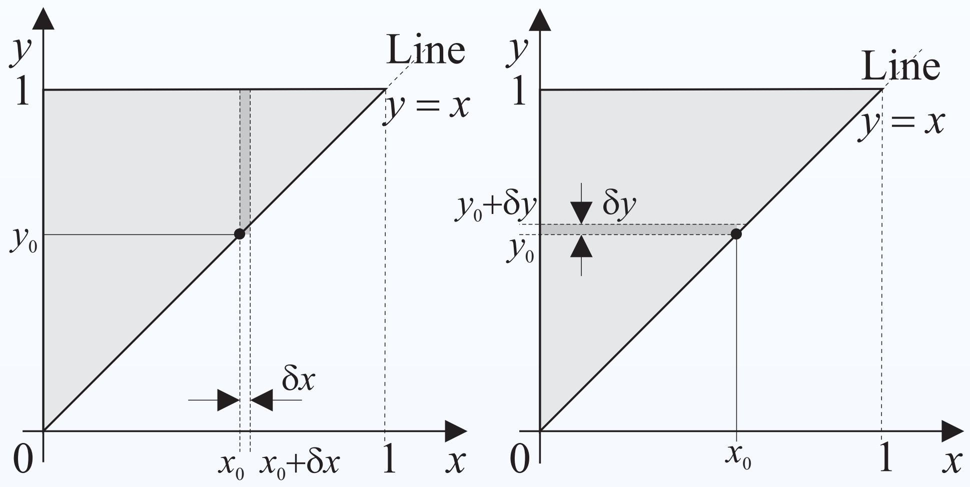
Example (Mean Vector). Let $f_{XY}(x, y) = 2$ for $0 < x < y < 1$ and zero otherwise. Find the mean-vector.

SOLUTION. Using the region-of-support:

$$f_X(x) = \int_{y=x}^{1} f_{XY}(x, y)\, dy = \int_{x}^{1} 2\, dy = 2(1 - x)$$

Mean Vectors and Correlation Matrices

Mean vector The **mean vector** is the first-moment :



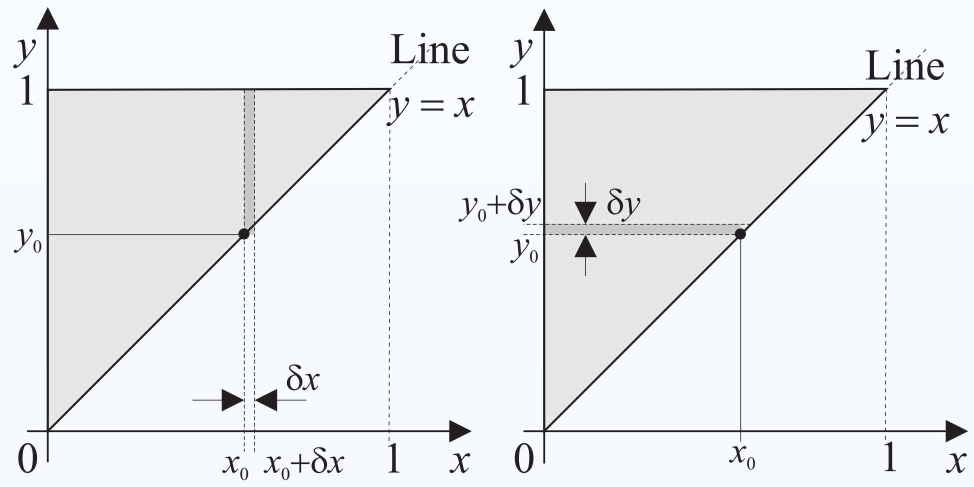
Example (Mean Vector). Let $f_{XY}(x, y) = 2$ for $0 < x < y < 1$ and zero otherwise. Find the mean-vector.

SOLUTION. Using the region-of-support:

$$f_Y(y) = \int_{x=0}^{y} f_{XY}(x, y)\, dx = \int_0^y 2\, dx = 2y$$

Mean Vectors and Correlation Matrices

Mean vector The **mean vector** is the first-moment :



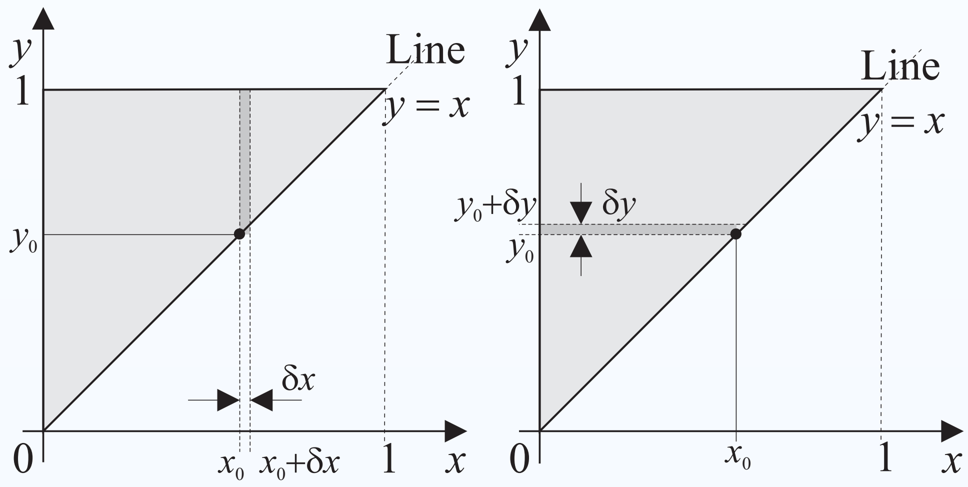
Example (Mean Vector). Let $f_{XY}(x, y) = 2$ for $0 < x < y < 1$ and zero otherwise. Find the mean-vector.

SOLUTION. Taking expectations then gives:

$$\mu_X = \int_0^1 x\, f_X(x)\, dx = \int_0^1 2x(1-x)\, dx$$

Mean Vectors and Correlation Matrices

Mean vector The **mean vector** is the first-moment :



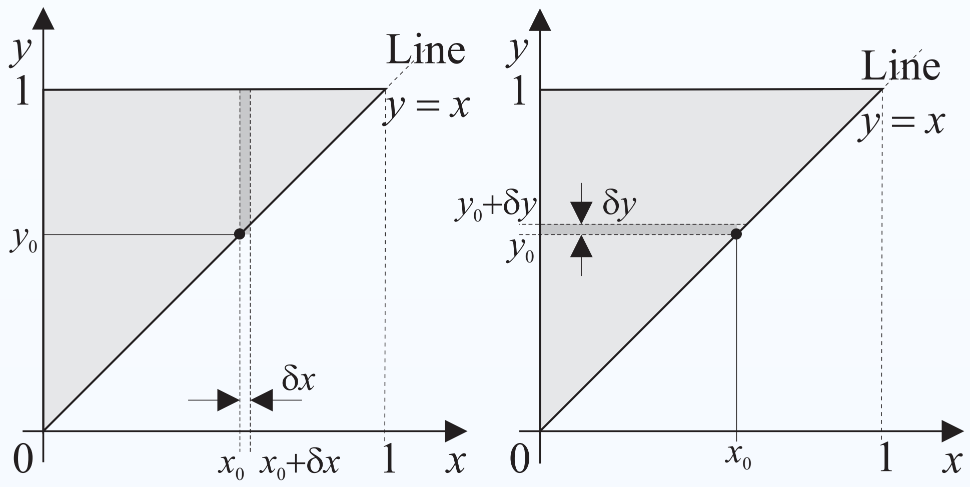
Example (Mean Vector). Let $f_{XY}(x, y) = 2$ for $0 < x < y < 1$ and zero otherwise. Find the mean-vector.

SOLUTION. Taking expectations then gives:

$$\mu_X = 2\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^1 = \frac{1}{3}$$

Mean Vectors and Correlation Matrices

Mean vector The **mean vector** is the first-moment :



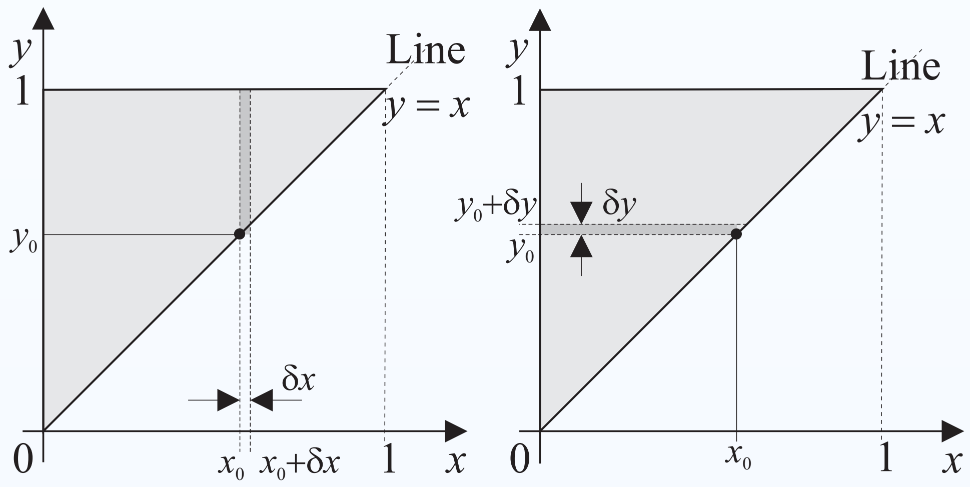
Example (Mean Vector). Let $f_{XY}(x, y) = 2$ for $0 < x < y < 1$ and zero otherwise. Find the mean-vector.

SOLUTION. Taking expectations then gives:

$$\mu_Y = \int_0^1 y\, f_Y(y)\, dy = 2\int_0^1 y^2\, dy = 2\left[\frac{y^3}{3}\right]_0^1 = \frac{2}{3} \qquad \square$$

Mean Vectors and Correlation Matrices

Mean vector The **mean vector** is the first-moment :



Example (Mean Vector). Let $f_{XY}(x, y) = 2$ for $0 < x < y < 1$ and zero otherwise. Find the mean-vector.

SOLUTION.

Mean Vectors and Correlation Matrices

Correlation Matrix The second-order moments of the random vector describe the spread of the distribution. The **autocorrelation matrix** is defined by:

$$\mathbf{R_X} \triangleq \mathbb{E}\left[\mathbf{X}(\zeta)\,\mathbf{X}^H(\zeta)\right] = \begin{bmatrix} r_{X_1X_1} & \cdots & r_{X_1X_N} \\ \vdots & \ddots & \vdots \\ r_{X_NX_1} & \cdots & r_{X_NX_N} \end{bmatrix}$$

Mean Vectors and Correlation Matrices

Correlation Matrix The second-order moments of the random vector describe the spread of the distribution. The **autocorrelation matrix** is defined by:

$$\mathbf{R_X} \triangleq \mathbb{E}\left[\mathbf{X}(\zeta)\,\mathbf{X}^H(\zeta)\right] = \begin{bmatrix} r_{X_1X_1} & \cdots & r_{X_1X_N} \\ \vdots & \ddots & \vdots \\ r_{X_NX_1} & \cdots & r_{X_NX_N} \end{bmatrix}$$

- The diagonal terms

$$r_{X_iX_i} \triangleq \mathbb{E}\left[|X_i(\zeta)|^2\right], \quad i \in \{1, \dots, N\}$$

are the second-order moments of each of the RVs, $X_i(\zeta)$.

Mean Vectors and Correlation Matrices

Correlation Matrix The second-order moments of the random vector describe the spread of the distribution. The **autocorrelation matrix** is defined by:

$$\mathbf{R_X} \triangleq \mathbb{E}\left[\mathbf{X}(\zeta)\,\mathbf{X}^H(\zeta)\right] = \begin{bmatrix} r_{X_1X_1} & \cdots & r_{X_1X_N} \\ \vdots & \ddots & \vdots \\ r_{X_NX_1} & \cdots & r_{X_NX_N} \end{bmatrix}$$

- The off-diagonal terms

$$r_{X_iX_j} \triangleq \mathbb{E}\left[X_i(\zeta)X_j^*(\zeta)\right] = r^*_{X_jX_i}, \quad i \neq j$$

measure the **correlation**, or statistical similarity, between RVs $X_i(\zeta)$ and $X_j(\zeta)$.

Mean Vectors and Correlation Matrices

Correlation Matrix The second-order moments of the random vector describe the spread of the distribution. The **autocorrelation matrix** is defined by:

$$\mathbf{R_X} \triangleq \mathbb{E}\left[\mathbf{X}(\zeta)\,\mathbf{X}^H(\zeta)\right] = \begin{bmatrix} r_{X_1 X_1} & \cdots & r_{X_1 X_N} \\ \vdots & \ddots & \vdots \\ r_{X_N X_1} & \cdots & r_{X_N X_N} \end{bmatrix}$$

- The off-diagonal terms

$$r_{X_i X_j} \triangleq \mathbb{E}\left[X_i(\zeta) X_j^*(\zeta)\right] = r^*_{X_j X_i}, \quad i \neq j$$

measure the **correlation** between RVs $X_i(\zeta)$ and $X_j(\zeta)$.

If $X_i(\zeta)$ and $X_j(\zeta)$ are **orthogonal**, their **correlation** is zero:

$$r_{X_i X_j} = \mathbb{E}\left[X_i(\zeta) X_j^*(\zeta)\right] = 0, \quad i \neq j$$

Mean Vectors and Correlation Matrices

Correlation Matrix Example (Correlation Matrix). Find the correlation matrix for random variables with joint-pdf given by $f_{XY}(x, y) = 2$ for $0 < x < y < 1$ and zero otherwise.

Mean Vectors and Correlation Matrices

Correlation Matrix Example (Correlation Matrix). Find the correlation matrix for random variables with joint-pdf given by $f_{XY}(x, y) = 2$ for $0 < x < y < 1$ and zero otherwise.

SOLUTION. The second-moments can utilise the marginals such that:

$$\mathbb{E}\left[X^2(\zeta)\right] = \int_0^1 x^2 f_X(x)\, dx = \int_0^1 2x^2(1-x)\, dx$$

$$= 2\left[\frac{x^3}{3} - \frac{x^4}{4}\right]_0^1 = \frac{1}{6}$$

$$\mathbb{E}\left[Y^2(\zeta)\right] = \int_0^1 y^2 f_Y(y)\, dy = 2\int_0^1 y^3\, dy = 2\left[\frac{y^4}{4}\right]_0^1 = \frac{1}{2}$$

□

Mean Vectors and Correlation Matrices

Correlation Matrix **Example (Correlation Matrix).** Find the correlation matrix for random variables with joint-pdf given by $f_{XY}(x, y) = 2$ for $0 < x < y < 1$ and zero otherwise.

SOLUTION. The correlation terms are given by:

$$\mathbb{E}\left[X(\zeta)\,Y(\zeta)\right] = \int_0^1 \int_0^y xy\, f_X Y\left(xy\right)\, dx\, dy$$

$$2\int_0^1 y \int_0^y x\, dx\, dy = 2\int_0^y y\left[\frac{x^2}{2}\right]_0^y dy$$

$$= \int_0^1 y^3\, dy = \left[\frac{y^4}{4}\right]_0^1 = \frac{1}{4}$$

□

Mean Vectors and Correlation Matrices

Correlation Matrix Example (Correlation Matrix). Find the correlation matrix for random variables with joint-pdf given by $f_{XY}(x, y) = 2$ for $0 < x < y < 1$ and zero otherwise.

SOLUTION. The correlation terms are given by:

$$\mathbb{E}[X(\zeta)\,Y(\zeta)] = \int_0^1 \int_0^y xy\, f_X Y(xy)\; dx\, dy$$

$$2\int_0^1 y \int_0^y x\, dx\, dy = 2\int_0^y y \left[\frac{x^2}{2}\right]_0^y dy$$

$$= \int_0^1 y^3\, dy = \left[\frac{y^4}{4}\right]_0^1 = \frac{1}{4} \qquad \square$$

This correlation matrix can be evaluated by the MATLAB expression:

Mean Vectors and Correlation Matrices

Correlation Matrix **Example (Correlation Matrix).** Find the correlation matrix for random variables with joint-pdf given by $f_{XY}(x, y) = 2$ for $0 < x < y < 1$ and zero otherwise.

SOLUTION. Hence, putting all of these calculations together gives the correlation matrix:

$$\mathbf{R}_{XY} = \begin{bmatrix} r_{XX} & r_{XY} \\ r_{YX} & r_{YY} \end{bmatrix} = \begin{bmatrix} \frac{1}{6} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} \end{bmatrix} \qquad \square$$

Mean Vectors and Correlation Matrices

– End-of-Topic 31: **Key Statistical definitions** –

Any Questions?

Properties of Correlation Matrices

It should be noticed that the **correlation** matrix is positive semidefinite; that is, the correlation matrices satisfies:

$$\mathbf{a}^H \mathbf{R}_{\mathbf{X}} \mathbf{a} \geq 0$$

for any complex vector $\mathbf{a}$.

Properties of Correlation Matrices

It should be noticed that the **correlation** matrix is positive semidefinite; that is, the correlation matrices satisfies:

$$\mathbf{a}^H \mathbf{R}_{\mathbf{X}} \mathbf{a} \geq 0$$

for any complex vector $\mathbf{a}$.

- This follows since:

$$\mathbf{a}^H \mathbf{R}_{\mathbf{X}} \mathbf{a} = \mathbf{a}^H \mathbb{E}\left[\mathbf{x}\mathbf{x}^H\right] \mathbf{a} = \mathbb{E}\left[\left|\mathbf{x}^H \mathbf{a}\right|^2\right]$$

Properties of Correlation Matrices

Theorem (Positive semi-definiteness). PROOF. Consider:

$$Y(\zeta) = \sum_{n=1}^{N} a_n X_n(\zeta) = \mathbf{a}^T \mathbf{X}(\zeta) \qquad \square$$

where $\mathbf{X}(\zeta) = \begin{bmatrix} X_1(\zeta) & \cdots & X_N(\zeta) \end{bmatrix}$ and $\mathbf{a} = \begin{bmatrix} \mathbf{a_1} & \cdots & \mathbf{a_N} \end{bmatrix}$ is an arbitrary vector of coefficients.

Properties of Correlation Matrices

Theorem (Positive semi-definiteness). PROOF. Consider:

$$Y(\zeta) = \sum_{n=1}^{N} a_n X_n(\zeta) = \mathbf{a}^T \mathbf{X}(\zeta)$$

where $\mathbf{X}(\zeta) = \begin{bmatrix} X_1(\zeta) & \cdots & X_N(\zeta) \end{bmatrix}$ and $\mathbf{a} = \begin{bmatrix} \mathbf{a_1} & \cdots & \mathbf{a_N} \end{bmatrix}$ is an arbitrary vector of coefficients.

The variance of $Y(\zeta)$ must, by definition, be positive, as must its second moment. Considering the second moment, then:

$$r_Y^{(2)} = \mathbb{E}\left[Y^2(\zeta)\right] = \mathbb{E}\left[\underbrace{\mathbf{a}^T \mathbf{X}(\zeta)\, \mathbf{X}(\zeta)^T \mathbf{a}}_{(1\times N)(N\times 1)(1\times N)(N\times 1)}\right]$$

$$= \mathbf{a}^T \mathbb{E}\left[\mathbf{X}(\zeta)\, \mathbf{X}(\zeta)^T\right] \mathbf{a} = \mathbf{a}^T \mathbf{R}_X\, \mathbf{a} \geq 0 \qquad \square$$

Properties of Correlation Matrices

Example (Valid correlation matrix). Determine whether the following is a valid correlation matrix:

$$\mathbf{R}_X = \begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix}$$

$\bowtie$

Properties of Correlation Matrices

Example (Valid correlation matrix). Determine whether the following is a valid correlation matrix:

$$\mathbf{R}_X = \begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix}$$

SOLUTION. This is not a valid correlation matrix as it is not symmetric, which is a requirement of a valid correlation matrix. In otherwords, $\mathbf{R}_X^T \neq \mathbf{R}_X$.

Properties of Correlation Matrices

Example (Valid correlation matrix). Determine whether the following is a valid correlation matrix:

$$\mathbf{R}_X = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

$\bowtie$

Properties of Correlation Matrices

Example (Valid correlation matrix). Determine whether the following is a valid correlation matrix:

$$\mathbf{R}_X = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

SOLUTION. Writing out the product $I = \mathbf{a}^T \mathbf{R}_X \mathbf{a}$ gives:

$$\begin{aligned} I &= \begin{bmatrix} \alpha & \beta \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \\ &= \begin{bmatrix} \alpha & \beta \end{bmatrix} \begin{bmatrix} \alpha + 2\beta \\ 2\alpha + \beta \end{bmatrix} \\ &= \alpha\left(\alpha + 2\beta\right) + \beta\left(2\alpha + \beta\right) \\ &= \underbrace{\alpha^2 + 4\alpha\beta + \beta^2}_{\text{look to complete the square}} \end{aligned}$$

Properties of Correlation Matrices

Example (Valid correlation matrix). Determine whether the following is a valid correlation matrix:

$$\mathbf{R}_X = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

SOLUTION. Writing out the product $I = \mathbf{a}^T \mathbf{R}_X \mathbf{a}$ gives:

$$I == \underbrace{\alpha^2 + 2\alpha\beta + \beta^2}_{\text{complete the square}} + 2\alpha\beta$$

$$= \underbrace{(\alpha+\beta)^2}_{\text{always positive}} + 2\alpha\beta \qquad \square$$

Noting the term $2\alpha\beta$ is not always positive, then selecting $\alpha = -\beta$, it follows that $I = -2\alpha^2 < 0$. Hence, $\mathbf{R}_X$ is not correlation matrix.

Properties of Correlation Matrices

– End-of-Topic 32: **Positive Semi-Definiteness for Correlation Matrices** –

Any Questions?

Further Statistical Descriptions

Covariance Matrix The **autocovariance matrix** is defined by:

$$\Gamma_{\mathbf{X}} \triangleq \mathbb{E}\left[\left(\mathbf{X}(\zeta) - \boldsymbol{\mu}_{\mathbf{X}}\right)\left(\mathbf{X}(\zeta) - \boldsymbol{\mu}_{\mathbf{X}}\right)^H\right] = \begin{bmatrix} \gamma_{X_1 X_1} & \cdots & \gamma_{X_1 X_N} \\ \vdots & \ddots & \cdots \\ \gamma_{X_N X_1} & \cdots & \gamma_{X_N X_N} \end{bmatrix}$$

Further Statistical Descriptions

Covariance Matrix The **autocovariance matrix** is defined by:

$$\mathbf{\Gamma}_{\mathbf{X}} \triangleq \mathbb{E}\left[\left(\mathbf{X}(\zeta) - \boldsymbol{\mu}_{\mathbf{X}}\right)\left(\mathbf{X}(\zeta) - \boldsymbol{\mu}_{\mathbf{X}}\right)^H\right] = \begin{bmatrix} \gamma_{X_1 X_1} & \cdots & \gamma_{X_1 X_N} \\ \vdots & \ddots & \cdots \\ \gamma_{X_N X_1} & \cdots & \gamma_{X_N X_N} \end{bmatrix}$$

- The diagonal terms

$$\gamma_{X_i X_i} \triangleq \sigma_{X_i}^2 = \mathbb{E}\left[\left|X_i(\zeta) - \mu_{X_i}\right|^2\right], \quad i \in \{1, \dots, N\}$$

are the **variances** of each of the RVs, $X_i(\zeta)$.

Further Statistical Descriptions

Covariance Matrix The **autocovariance matrix** is defined by:

$$\mathbf{\Gamma_X} \triangleq \mathbb{E}\left[\left(\mathbf{X}(\zeta) - \boldsymbol{\mu}_\mathbf{X}\right)\left(\mathbf{X}(\zeta) - \boldsymbol{\mu}_\mathbf{X}\right)^H\right] = \begin{bmatrix} \gamma_{X_1 X_1} & \cdots & \gamma_{X_1 X_N} \\ \vdots & \ddots & \cdots \\ \gamma_{X_N X_1} & \cdots & \gamma_{X_N X_N} \end{bmatrix}$$

- The off-diagonal terms

$$\gamma_{X_i X_j} \triangleq \mathbb{E}\left[\left(X_i(\zeta) - \mu_{X_i}\right)\left(X_j(\zeta) - \mu_{X_j}\right)^*\right]$$
$$= r_{X_i X_j} - \mu_{X_i}\mu^*_{X_j} = \gamma^*_{X_j X_i}, \quad i \neq j$$

measure the **covariance** $X_i(\zeta)$ and $X_j(\zeta)$.

Further Statistical Descriptions

Covariance Matrix The **autocovariance matrix** is defined by:

$$\mathbf{\Gamma}_{\mathbf{X}} \triangleq \mathbb{E}\left[\left(\mathbf{X}\left(\zeta\right)-\boldsymbol{\mu}_{\mathbf{X}}\right)\left(\mathbf{X}\left(\zeta\right)-\boldsymbol{\mu}_{\mathbf{X}}\right)^H\right] = \begin{bmatrix} \gamma_{X_1 X_1} & \cdots & \gamma_{X_1 X_N} \\ \vdots & \ddots & \cdots \\ \gamma_{X_N X_1} & \cdots & \gamma_{X_N X_N} \end{bmatrix}$$

It can easily be shown that the **covariance** matrix, $\mathbf{\Gamma}_{\mathbf{X}}$, must also be positive-semi definite, and is also a Hermitian matrix.

$$\mathbf{a}^H \mathbf{\Gamma}_{\mathbf{X}} \mathbf{a} \geq 0$$

Further Statistical Descriptions

- Moreover, as for scalar RVs, the covariance, $\gamma_{X_i X_j}$, can be expressed in terms of the standard deviations of $X_i(\zeta)$ and $X_j(\zeta)$:

$$\rho_{X_i X_j} \triangleq \frac{\gamma_{X_i X_j}}{\sigma_{X_i} \sigma_{X_j}} = \rho^*_{X_j X_i}$$

Further Statistical Descriptions

- Moreover, as for scalar RVs, the covariance, $\gamma_{X_i X_j}$, can be expressed in terms of the standard deviations of $X_i(\zeta)$ and $X_j(\zeta)$:

$$\rho_{X_i X_j} \triangleq \frac{\gamma_{X_i X_j}}{\sigma_{X_i} \sigma_{X_j}} = \rho^*_{X_j X_i}$$

- Again, the correlation coefficient measures the degree of statistical similarity between two random variables.

Note that:

- If $\left|\rho_{X_i X_j}\right| = 1$, $i \neq j$, then the RVs are said to be *perfectly correlated*.
- However, if $\rho_{X_i X_j} = 0$, which occurs when the covariance $\gamma_{X_i X_j} = 0$, then the RVs are said to be *uncorrelated*.

Further Statistical Descriptions

The autocorrelation and autocovariance matrices are related, and it can easily be seen that:

$$\mathbf{\Gamma}_{\mathbf{X}} \triangleq \mathbb{E}\left[\left[\mathbf{X}(\zeta) - \boldsymbol{\mu}_{\mathbf{X}}\right] \left[\mathbf{X}(\zeta) - \boldsymbol{\mu}_{\mathbf{X}}\right]^H \right] = \mathbf{R}_{\mathbf{X}} - \boldsymbol{\mu}_{\mathbf{X}} \boldsymbol{\mu}_{\mathbf{X}}^H$$

Further Statistical Descriptions

The autocorrelation and autocovariance matrices are related, and it can easily be seen that:

$$\mathbf{\Gamma_X} \triangleq \mathbb{E}\left[\left[\mathbf{X}(\zeta) - \boldsymbol{\mu}_\mathbf{X}\right] \left[\mathbf{X}(\zeta) - \boldsymbol{\mu}_\mathbf{X}\right]^H \right] = \mathbf{R_X} - \boldsymbol{\mu}_\mathbf{X} \boldsymbol{\mu}_\mathbf{X}^H$$

In fact, if $\boldsymbol{\mu}_\mathbf{X} = 0$, then $\mathbf{\Gamma_X} = \mathbf{R_X}$.

If the random variables $X_i(\zeta)$ and $X_j(\zeta)$ are **independent**, then they are also **uncorrelated** since:

$$\begin{aligned} r_{X_i X_j} &= \mathbb{E}\left[X_i(\zeta)\, X_j(\zeta)^*\right] = \mathbb{E}\left[X_i(\zeta)\right] \mathbb{E}\left[X_j^*(\zeta)\right] \\ &= \mu_{X_i} \mu_{X_j}^* \quad \Rightarrow \quad \gamma_{X_i X_j} = 0 \end{aligned}$$

Further Statistical Descriptions

The autocorrelation and autocovariance matrices are related, and it can easily be seen that:

$$\mathbf{\Gamma_X} \triangleq \mathbb{E}\left[\left[\mathbf{X}(\zeta) - \boldsymbol{\mu}_\mathbf{X}\right]\left[\mathbf{X}(\zeta) - \boldsymbol{\mu}_\mathbf{X}\right]^H\right] = \mathbf{R_X} - \boldsymbol{\mu}_\mathbf{X}\boldsymbol{\mu}_\mathbf{X}^H$$

In fact, if $\boldsymbol{\mu}_\mathbf{X} = 0$, then $\mathbf{\Gamma_X} = \mathbf{R_X}$.

If the random variables $X_i(\zeta)$ and $X_j(\zeta)$ are **independent**, then they are also **uncorrelated** since:

$$\begin{aligned} r_{X_i X_j} &= \mathbb{E}\left[X_i(\zeta)\, X_j(\zeta)^*\right] = \mathbb{E}\left[X_i(\zeta)\right]\mathbb{E}\left[X_j^*(\zeta)\right] \\ &= \mu_{X_i}\mu_{X_j}^* \quad \Rightarrow \quad \gamma_{X_i X_j} = 0 \end{aligned}$$

- Note, however, that uncorrelatedness does not imply independence, unless the RVs are jointly-Gaussian.

Further Statistical Descriptions

Cross-correlation is defined as

$$
\mathbf{R}_{\mathbf{XY}} \triangleq \mathbb{E}\left[\mathbf{X}(\zeta)\,\mathbf{Y}^H(\zeta)\right]
= \begin{bmatrix} \mathbb{E}\left[X_1(\zeta)Y_1^*(\zeta)\right] & \cdots & \mathbb{E}\left[X_1(\zeta)Y_M^*(\zeta)\right] \\ \vdots & \ddots & \vdots \\ \mathbb{E}\left[X_N(\zeta)Y_1^*(\zeta)\right] & \cdots & \mathbb{E}\left[X_N(\zeta)Y_M^*(\zeta)\right] \end{bmatrix}
$$

Further Statistical Descriptions

Cross-correlation is defined as

$$\mathbf{R_{XY}} \triangleq \mathbb{E}\left[\mathbf{X}(\zeta)\,\mathbf{Y}^H(\zeta)\right] = \begin{bmatrix} \mathbb{E}\left[X_1(\zeta)Y_1^*(\zeta)\right] & \cdots & \mathbb{E}\left[X_1(\zeta)Y_M^*(\zeta)\right] \\ \vdots & \ddots & \vdots \\ \mathbb{E}\left[X_N(\zeta)Y_1^*(\zeta)\right] & \cdots & \mathbb{E}\left[X_N(\zeta)Y_M^*(\zeta)\right] \end{bmatrix}$$

Cross-covariance is defined as

$$\mathbf{\Gamma_{XY}} \triangleq \mathbb{E}\left[\{\mathbf{X}(\zeta) - \boldsymbol{\mu}_\mathbf{X}\}\{\mathbf{Y}(\zeta) - \boldsymbol{\mu}_\mathbf{Y}\}^H\right] = \mathbf{R_{XY}} - \boldsymbol{\mu}_\mathbf{X}\boldsymbol{\mu}_\mathbf{Y}^H$$

Further Statistical Descriptions

Cross-correlation is defined as

$$\mathbf{R_{XY}} \triangleq \mathbb{E}\left[\mathbf{X}(\zeta)\,\mathbf{Y}^H(\zeta)\right] = \begin{bmatrix} \mathbb{E}\left[X_1(\zeta)Y_1^*(\zeta)\right] & \cdots & \mathbb{E}\left[X_1(\zeta)Y_M^*(\zeta)\right] \\ \vdots & \ddots & \vdots \\ \mathbb{E}\left[X_N(\zeta)Y_1^*(\zeta)\right] & \cdots & \mathbb{E}\left[X_N(\zeta)Y_M^*(\zeta)\right] \end{bmatrix}$$

Cross-covariance is defined as

$$\mathbf{\Gamma_{XY}} \triangleq \mathbb{E}\left[\{\mathbf{X}(\zeta) - \boldsymbol{\mu}_\mathbf{X}\}\{\mathbf{Y}(\zeta) - \boldsymbol{\mu}_\mathbf{Y}\}^H\right] = \mathbf{R_{XY}} - \boldsymbol{\mu}_\mathbf{X}\boldsymbol{\mu}_\mathbf{Y}^H$$

- Uncorrelated if $\mathbf{\Gamma_{XY}} = 0 \quad \Rightarrow \quad \mathbf{R_{XY}} = \boldsymbol{\mu}_\mathbf{X}\boldsymbol{\mu}_\mathbf{Y}^H$.
- Orthogonal if $\mathbf{R_{XY}} = 0$.

Further Statistical Descriptions

Example (Sum of Random Vectors). Consider the sum of two zero-mean random vectors that are uncorrelated. What are the correlation and covariance matrices of the sum?

Further Statistical Descriptions

Example (Sum of Random Vectors). Consider the sum of two zero-mean random vectors that are uncorrelated. What are the correlation and covariance matrices of the sum?

SOLUTION. Let $\mathbf{Z}(\zeta) = \mathbf{X}(\zeta) + \mathbf{Y}(\zeta)$. Then:

$$\begin{aligned}
\mathbf{R}_{\mathbf{Z}} &= \mathbb{E}\left[\mathbf{Z}(\zeta)\ \mathbf{Z}^H(\zeta)\right] = \mathbb{E}\left[(\mathbf{X}(\zeta) + \mathbf{Y}(\zeta))\left(\mathbf{X}(\zeta) + \mathbf{Y}(\zeta)\right)^H\right] \\
&= \mathbb{E}\left[\mathbf{X}(\zeta)\,\mathbf{X}^H(\zeta)\right] + \mathbb{E}\left[\mathbf{X}(\zeta)\,\mathbf{Y}^H(\zeta)\right] \\
&\qquad + \mathbb{E}\left[\mathbf{Y}(\zeta)\,\mathbf{X}^H(\zeta)\right] + \mathbb{E}\left[\mathbf{Y}(\zeta)\,\mathbf{Y}^H(\zeta)\right] \\
&= \mathbf{R}_{\mathbf{X}} + \mathbf{R}_{\mathbf{XY}} + \mathbf{R}_{\mathbf{YX}} + \mathbf{R}_{\mathbf{YY}}
\end{aligned}$$

□

Further Statistical Descriptions

Example (Sum of Random Vectors). Consider the sum of two zero-mean random vectors that are uncorrelated. What are the correlation and covariance matrices of the sum?

SOLUTION. Let $\mathbf{Z}(\zeta) = \mathbf{X}(\zeta) + \mathbf{Y}(\zeta)$. Then:

$$\begin{aligned}
\mathbf{R}_\mathbf{Z} &= \mathbb{E}\left[\mathbf{Z}(\zeta)\ \mathbf{Z}^H(\zeta)\right] = \mathbb{E}\left[(\mathbf{X}(\zeta) + \mathbf{Y}(\zeta))\left(\mathbf{X}(\zeta) + \mathbf{Y}(\zeta)\right)^H\right] \\
&= \mathbb{E}\left[\mathbf{X}(\zeta)\,\mathbf{X}^H(\zeta)\right] + \mathbb{E}\left[\mathbf{X}(\zeta)\,\mathbf{Y}^H(\zeta)\right] \\
&\qquad + \mathbb{E}\left[\mathbf{Y}(\zeta)\,\mathbf{X}^H(\zeta)\right] + \mathbb{E}\left[\mathbf{Y}(\zeta)\,\mathbf{Y}^H(\zeta)\right] \\
&= \mathbf{R}_\mathbf{X} + \mathbf{R}_\mathbf{XY} + \mathbf{R}_\mathbf{YX} + \mathbf{R}_\mathbf{YY}
\end{aligned}$$

- Since the random vectors are uncorrelated, then $\mathbf{R}_\mathbf{XY} = \mathbf{R}_\mathbf{YX} = \mathbf{0}$, and therefore $\mathbf{R}_\mathbf{Z} = \mathbf{R}_\mathbf{X} + \mathbf{R}_\mathbf{Y}$.
- Moreover, the covariance matrix is equal to the correlation matrix as the random vectors are zero-mean. □

Further Statistical Descriptions

– End-of-Topic 33: **Further Statistical Descriptions** –

Any Questions?

Multivariate Gaussian Density Function

Gaussian random vectors play a very important role in the design and analysis of signal processing systems. A Gaussian random vector is characterised by a multivariate Normal density.

For a *real* random vector, this density function has the form:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{N}{2}} \left|\mathbf{\Gamma}_{\mathbf{X}}\right|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{X}})^T \mathbf{\Gamma}_{\mathbf{X}}^{-1} (\mathbf{x}-\boldsymbol{\mu}_{\mathbf{X}})\right]$$

Multivariate Gaussian Density Function

Gaussian random vectors play a very important role in the design and analysis of signal processing systems. A Gaussian random vector is characterised by a multivariate Normal density.

For a *real* random vector, this density function has the form:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{N}{2}} \left|\mathbf{\Gamma}_{\mathbf{X}}\right|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{X}})^T \mathbf{\Gamma}_{\mathbf{X}}^{-1} (\mathbf{x}-\boldsymbol{\mu}_{\mathbf{X}})\right]$$

where N is the dimension of $\mathbf{X}(\zeta)$, and $\mathbf{X}(\zeta)$ has mean $\boldsymbol{\mu}_{\mathbf{X}}$ and covariance $\mathbf{\Gamma}_{\mathbf{X}}$. It is often denoted as:

$$f_{\mathbf{X}}(\mathbf{x}) = \mathcal{N}\left(\mathbf{x} \,\middle|\, \boldsymbol{\mu}_{\mathbf{X}}, \mathbf{\Gamma}_{\mathbf{X}}\right)$$

Multivariate Gaussian Density Function

Gaussian random vectors play a very important role in the design and analysis of signal processing systems. A Gaussian random vector is characterised by a multivariate Normal density.

For a *real* random vector, this density function has the form:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{N}{2}} \left|\mathbf{\Gamma_X}\right|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{X}})^T \mathbf{\Gamma}_{\mathbf{X}}^{-1}(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{X}})\right]$$

where N is the dimension of $\mathbf{X}(\zeta)$, and $\mathbf{X}(\zeta)$ has mean $\boldsymbol{\mu}_{\mathbf{X}}$ and covariance $\mathbf{\Gamma}_{\mathbf{X}}$. It is often denoted as:

$$f_{\mathbf{X}}(\mathbf{x}) = \mathcal{N}\left(\mathbf{x} \,\middle|\, \boldsymbol{\mu}_{\mathbf{X}}, \mathbf{\Gamma}_{\mathbf{X}}\right)$$

The notation when a random vector is sampled from a normal:

$$\mathbf{X}(\zeta) \sim \mathcal{N}\left(\boldsymbol{\mu}_{\mathbf{X}}, \mathbf{\Gamma}_{\mathbf{X}}\right)$$

Deriving the Multivariate Gaussian

The pdf for the multivariate Gaussian is often quoted, but where does it come from?

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{N}{2}} \left|\mathbf{\Gamma}_{\mathbf{X}}\right|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{X}})^T \mathbf{\Gamma}_{\mathbf{X}}^{-1}(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{X}})\right]$$

Deriving the Multivariate Gaussian

The pdf for the multivariate Gaussian is often quoted, but where does it come from?

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{N}{2}} \left|\mathbf{\Gamma_X}\right|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{X}})^T \mathbf{\Gamma}_{\mathbf{X}}^{-1}(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{X}})\right]$$

- Suppose that NRVs, $X_n(\zeta)$ for $n \in \{0, \dots, N-1\}$, are independent zero-mean unit variance Gaussian densities, and each have pdf given by $f_{X_n}(x_n)$.
- Then the joint-pdf of $\mathbf{X}(\zeta) = [X_0(\zeta), \cdots, X_{N-1}(\zeta)]^T$ is:

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{n=0}^{N-1} f_{X_n}(x_n)$$

Deriving the Multivariate Gaussian

Since $X_n(\zeta)$ is Gaussian distributed:

$$f_{X_n}(x_n) = \mathcal{N}\left(x_n \,\middle|\, 0,\, 1\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x_n^2}{2}}$$

and hence it follows that:

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_n^2}{2}} = \frac{1}{(2\pi)^{\frac{N}{2}}} e^{-\frac{1}{2}\sum_{n=0}^{N-1} x_n^2}$$

Deriving the Multivariate Gaussian

Since $X_n(\zeta)$ is Gaussian distributed:

$$f_{X_n}(x_n) = \mathcal{N}\left(x_n \,\middle|\, 0,\, 1\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x_n^2}{2}}$$

and hence it follows that:

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_n^2}{2}} = \frac{1}{(2\pi)^{\frac{N}{2}}} e^{-\frac{1}{2}\sum_{n=0}^{N-1} x_n^2}$$

Defining the vector $\mathbf{x} = [x_0, \cdots, x_{N-1}]^T$, then it follows that

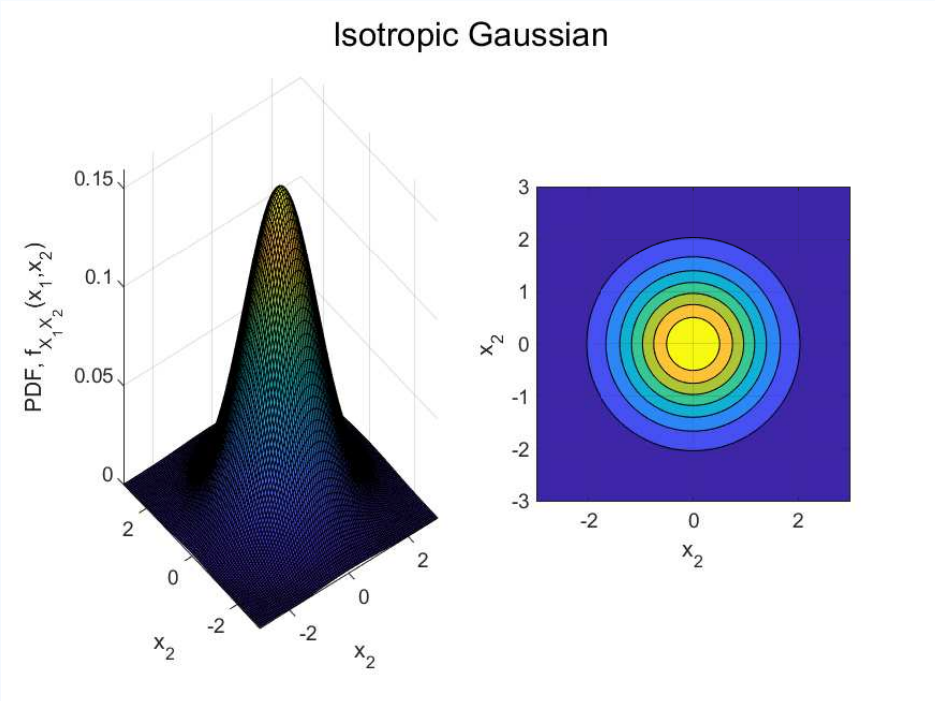
$$\mathbf{x}^T\mathbf{x} = \begin{bmatrix} x_1 & \cdots & x_N \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} = \sum_{n=0}^{N-1} x_n^2$$

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{N}{2}}} e^{-\frac{1}{2}\mathbf{x}^T\mathbf{x}}$$

Deriving the Multivariate Gaussian

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{N}{2}}} e^{-\frac{1}{2}\mathbf{x}^T \mathbf{x}}$$

This is an **isotropic Gaussian**, which is circularly symmetric.



A graphical representation of an isotropic Gaussian random vector.

Deriving the Multivariate Gaussian

A non-isotropic Gaussian can be obtained by a linear shift, scale, and rotation using the linear transformations. Hence, set:

$$\mathbf{y} = \mathbf{A}\,\mathbf{x} + \boldsymbol{\mu}$$

Deriving the Multivariate Gaussian

A non-isotropic Gaussian can be obtained by a linear shift, scale, and rotation using the linear transformations. Hence, set:

$$\mathbf{y} = \mathbf{A}\,\mathbf{x} + \boldsymbol{\mu}$$

Apply the probability transformation rule, noting one solution $\mathbf{x} = \mathbf{A}^{-1}\left(\mathbf{y} - \boldsymbol{\mu}\right)$ and Jacobian $J_{\mathbf{x}\to\mathbf{y}} = \det \mathbf{A}$

$$f_{\mathbf{Y}}\left(\mathbf{y}\right) = \frac{f_{\mathbf{X}}\left(\mathbf{A}^{-1}\left(\mathbf{y} - \boldsymbol{\mu}\right)\right)}{\left|\det \mathbf{A}\right|}$$

Deriving the Multivariate Gaussian

A non-isotropic Gaussian can be obtained by a linear shift, scale, and rotation using the linear transformations. Hence, set:

$$\mathbf{y} = \mathbf{A}\,\mathbf{x} + \boldsymbol{\mu}$$

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{|\det \mathbf{A}|} \frac{1}{(2\pi)^{\frac{N}{2}}} \exp\left[-\frac{1}{2}\left(\mathbf{A}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right)^T \left(\mathbf{A}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right)\right]$$

$$= \frac{1}{(2\pi)^{\frac{N}{2}} \left|\mathbf{A}^T\mathbf{A}\right|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^T \mathbf{A}^{-T}\mathbf{A}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right]$$

where it has been noted that $\left|\mathbf{A}\mathbf{A}^T\right|^{\frac{1}{2}} = \det \mathbf{A}$.

Deriving the Multivariate Gaussian

A non-isotropic Gaussian can be obtained by a linear shift, scale, and rotation using the linear transformations. Hence, set:

$$\mathbf{y} = \mathbf{A}\,\mathbf{x} + \boldsymbol{\mu}$$

Finally, writing $\boldsymbol{\Gamma}_{\mathbf{Y}} = \mathbf{A}\,\mathbf{A}^T$ and $\boldsymbol{\mu}_{\mathbf{Y}} = \boldsymbol{\mu}$, then:

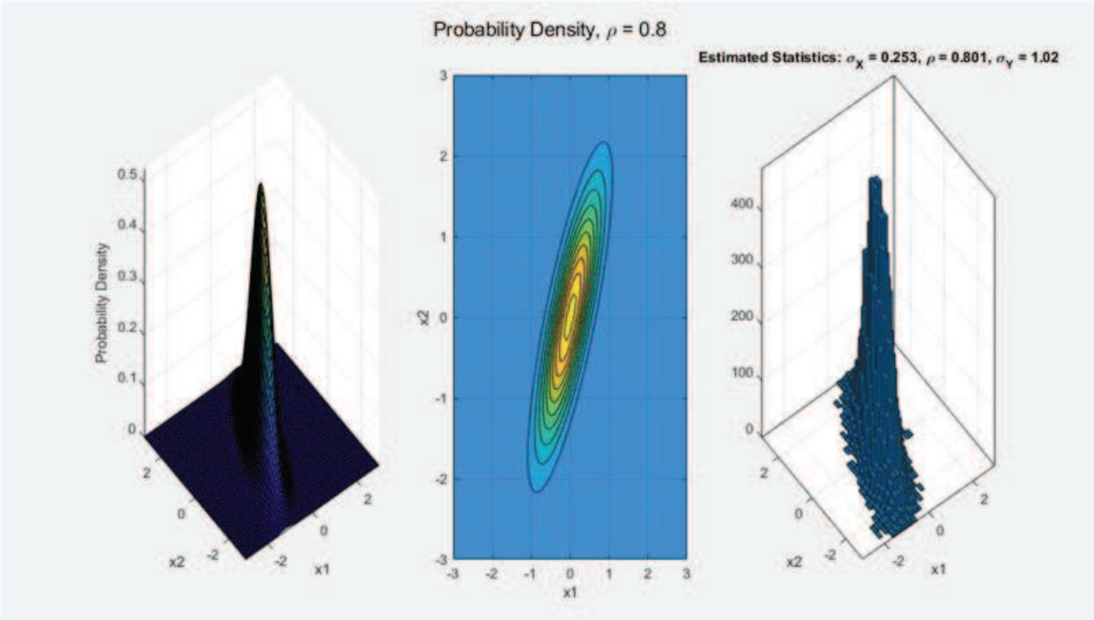
$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}) &= \frac{1}{(2\pi)^{\frac{N}{2}} \left|\boldsymbol{\Gamma}_{\mathbf{Y}}\right|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\left(\mathbf{y}-\boldsymbol{\mu}_{\mathbf{Y}}\right)^T \boldsymbol{\Gamma}_{\mathbf{Y}}^{-1} \left(\mathbf{y}-\boldsymbol{\mu}_{\mathbf{Y}}\right)\right] \\ &= \mathcal{N}\left(\mathbf{y} \,\middle|\, \boldsymbol{\mu}_{\mathbf{Y}}, \boldsymbol{\Gamma}_{\mathbf{Y}}\right) \end{aligned}$$

Deriving the Multivariate Gaussian

Using the definition of the correlation coefficient, for a bivariate Gaussian, the covariance matrix can be written as:

$$\mathbf{\Gamma_Y} = \begin{bmatrix} \sigma_{Y_1}^2 & \rho_{Y_1 Y_2}\sigma_{Y_1}\sigma_{Y_2} \\ \rho_{Y_1 Y_2}\sigma_{Y_1}\sigma_{Y_2} & \sigma_{Y_2}^2 \end{bmatrix}$$

The pdf can then be plotted as $\rho_{Y_1 Y_2}$ changes.



Properties of Multivariate Gaussians

The normal distribution is a useful model of a random vector because of its many important properties.

Properties of Multivariate Gaussians

The normal distribution is a useful model of a random vector because of its many important properties.

1. $f_{\mathbf{X}}(\mathbf{x}) = \mathcal{N}\left(\mathbf{x} \,\middle|\, \boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\Gamma}_{\mathbf{X}}\right)$ is completely specified by its mean $\boldsymbol{\mu}_{\mathbf{X}}$ and covariance $\boldsymbol{\Gamma}_{\mathbf{X}}$.

2. If the components of $\mathbf{X}(\zeta)$ are mutually uncorrelated, then they are also independent.

Properties of Multivariate Gaussians

The normal distribution is a useful model of a random vector because of its many important properties.

1. $f_{\mathbf{X}}(\mathbf{x}) = \mathcal{N}\left(\mathbf{x} \,\middle|\, \boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\Gamma}_{\mathbf{X}}\right)$ is completely specified by its mean $\boldsymbol{\mu}_{\mathbf{X}}$ and covariance $\boldsymbol{\Gamma}_{\mathbf{X}}$.

2. If the components of $\mathbf{X}(\zeta)$ are mutually uncorrelated, then they are also independent.

3. A linear transformation of a normal random vector is also normal.

   This is a particularly useful, since the output of a linear system subject to a Gaussian input is also Gaussian.

Properties of Multivariate Gaussians

The normal distribution is a useful model of a random vector because of its many important properties.

1. $f_{\mathbf{X}}(\mathbf{x}) = \mathcal{N}\left(\mathbf{x} \,\middle|\, \boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\Gamma}_{\mathbf{X}}\right)$ is completely specified by its mean $\boldsymbol{\mu}_{\mathbf{X}}$ and covariance $\boldsymbol{\Gamma}_{\mathbf{X}}$.

2. If the components of $\mathbf{X}(\zeta)$ are mutually uncorrelated, then they are also independent.

3. A linear transformation of a normal random vector is also normal.

   This is a particularly useful, since the output of a linear system subject to a Gaussian input is also Gaussian.

4. If $\mathbf{X}(\zeta)$ and $\mathbf{Y}(\zeta)$ are *jointly*-Gaussian, then so are their *marginal*-distributions, and their *conditional*-distributions.

Properties of Multivariate Gaussians

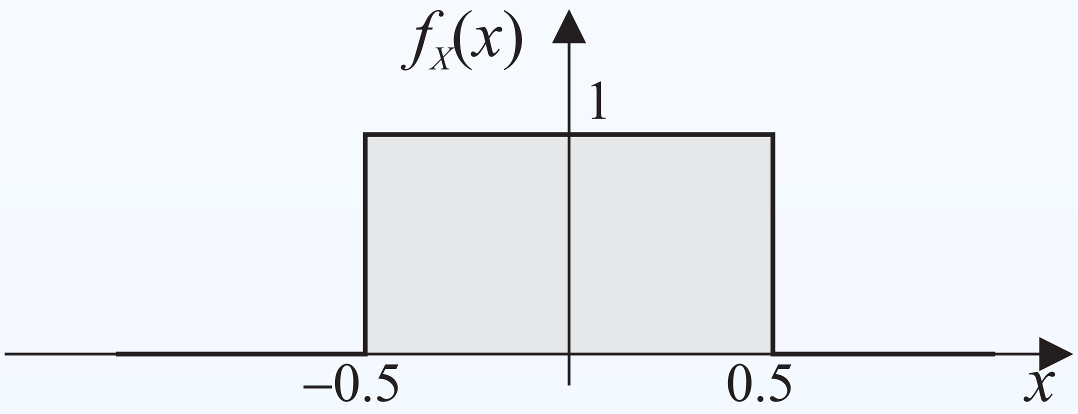
– End-of-Topic 34: **Multivariate Gaussian Distribution** –

Any Questions?

Central limit theorem

To motivate the central limit theorem, consider the following example.

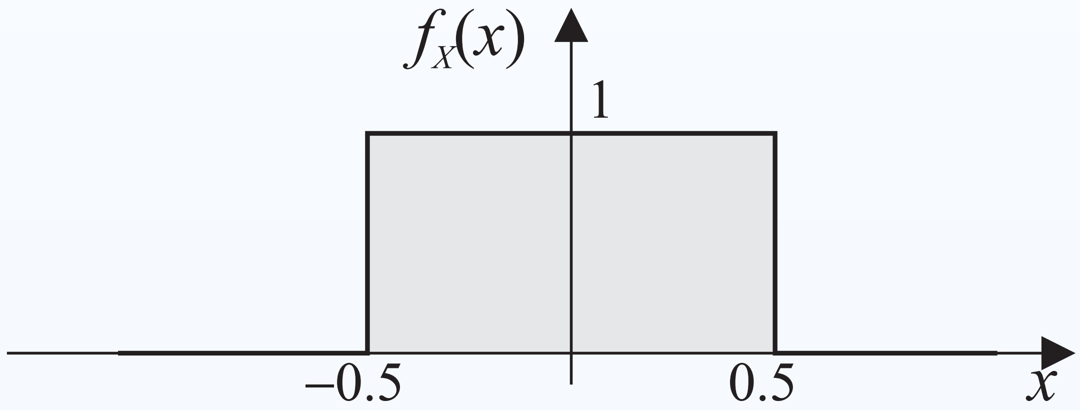
Example. Suppose $\{X_k(\zeta)\}_{k=1}^4$ are four i. i. d. random variables uniformally distributed over $[-0.5, 0, 5]$. Compute and plot the pdfs of $Y_M(\zeta) \triangleq \sum_{k=1}^M X_k(\zeta)$ for $M = \{2, 3, 4\}$.



⋈

Central limit theorem

Example. Suppose $\{X_k(\zeta)\}_{k=1}^4$ are four i. i. d. random variables uniformally distributed over $[-0.5, 0, 5]$. Compute and plot the pdfs of $Y_M(\zeta) \triangleq \sum_{k=1}^M X_k(\zeta)$ for $M = \{2, 3, 4\}$.



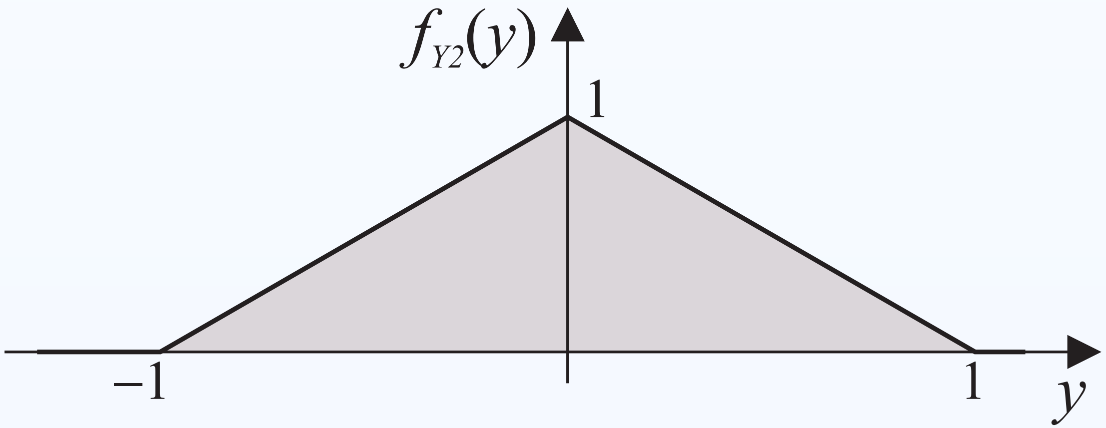
SOLUTION. Using the convolution result for the sum of independent random variables, it follows:

$$f_{Y_2}(y) = f_{X_1}(y) * f_{X_2}(y) = f_X(y) * f_X(y)$$
$$f_{Y_3}(y) = f_{Y_2}(y) * f_{X_3}(y) = f_{Y_2}(y) * f_X(y)$$
$$f_{Y_4}(y) = f_{Y_3}(y) * f_{X_4}(y) = f_{Y_3}(y) * f_X(y)$$

Central limit theorem

Example. Suppose $\{X_k(\zeta)\}_{k=1}^4$ are four i. i. d. random variables uniformally distributed over $[-0.5, 0, 5]$. Compute and plot the pdfs of $Y_M(\zeta) \triangleq \sum_{k=1}^M X_k(\zeta)$ for $M = \{2, 3, 4\}$.

SOLUTION. The convolution calculations:



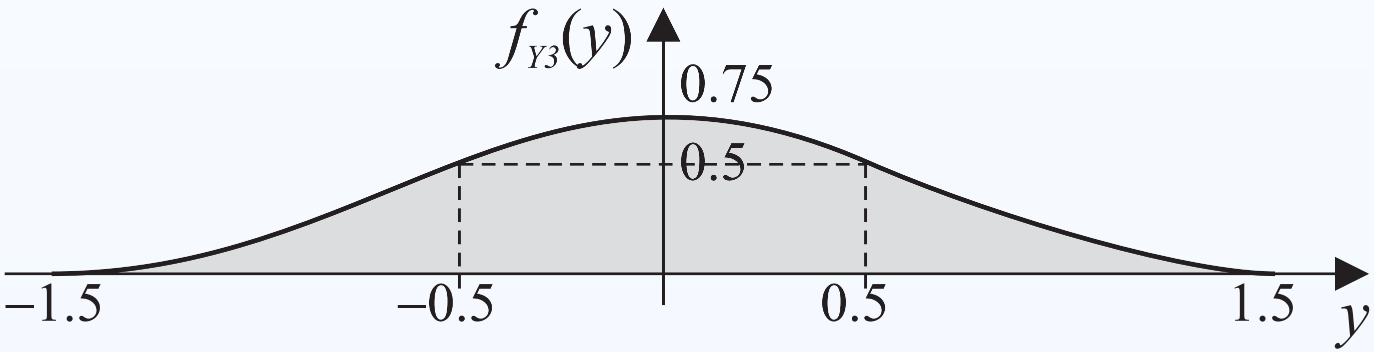
$$f_{Y_2}(y) = \begin{cases} 1 + y & -1 \leq y < 0 \\ 1 - y & 0 \leq y < 1 \\ 0 & \text{otherwise} \end{cases}$$

□

Central limit theorem

Example. Suppose $\{X_k(\zeta)\}_{k=1}^4$ are four i. i. d. random variables uniformally distributed over $[-0.5, 0, 5]$. Compute and plot the pdfs of $Y_M(\zeta) \triangleq \sum_{k=1}^M X_k(\zeta)$ for $M = \{2, 3, 4\}$.

SOLUTION. The convolution calculations:

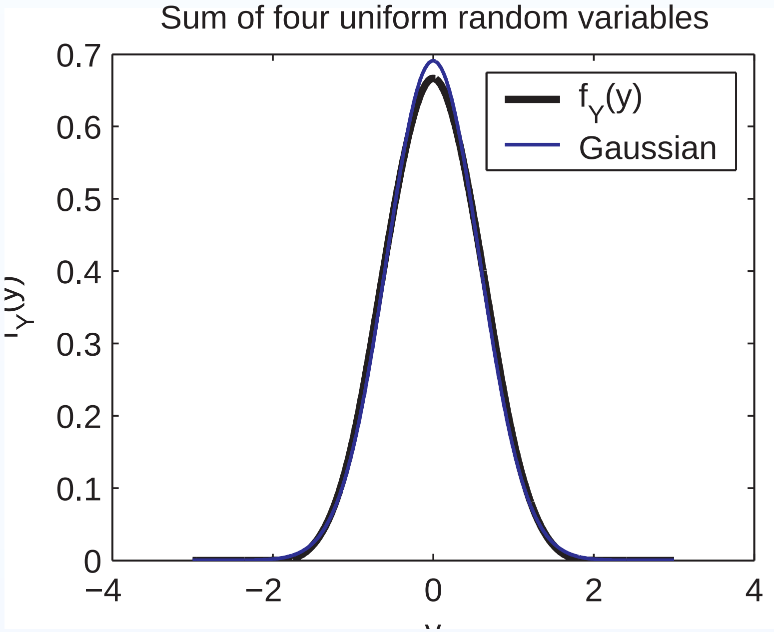


$$f_{Y_3}(y) = \begin{cases} \frac{1}{2}\left(y + \frac{3}{2}\right)^2 & -\frac{3}{2} \leq y < -\frac{1}{2} \\ \frac{3}{4} - y^2 & -\frac{1}{2} \leq y < \frac{1}{2} \\ \frac{1}{2}\left(y - \frac{3}{2}\right)^2 & \frac{1}{2} \leq y < \frac{3}{2} \\ 0 & \text{otherwise} \end{cases}$$

□

Central limit theorem

Example. SOLUTION. The convolution calculations:



The pdf of $f_{Y_4}(y)$, and also the pdf of $\mathcal{N}\left(y \,\middle|\, 0, \frac{1}{3}\right)$.

$$f_{Y_4}(y) = \begin{cases} \frac{1}{6}(y+2)^3 & -2 \le y < -1 \\ -\frac{1}{2}y^3 - y^2 + \frac{2}{3} & -1 \le y < 0 \\ \frac{1}{2}y^3 - y^2 + \frac{2}{3} & 0 \le y < 1 \\ -\frac{1}{6}(y-2)^3 & 1 \le y < 2 \\ 0 & \text{otherwise} \end{cases} \qquad \square$$

Central limit theorem

Consider the random variable $Y(\zeta)$ given by:

$$Y_M(\zeta) = \sum_{k=1}^{M} X_k(\zeta)$$

What is the distribution of $Y_M(\zeta)$ as $M \to \infty$?

Central limit theorem

Consider the random variable $Y(\zeta)$ given by:

$$Y_M(\zeta) = \sum_{k=1}^{M} X_k(\zeta)$$

What is the distribution of $Y_M(\zeta)$ as $M \to \infty$?

Informally, the CLT is well known, and the answer is a Gaussian. Assume that the $X_M(\zeta)$'s are i. i. d., and the mean and variance of $X_m(\zeta)$ are finite and given by μ_X and σ_X^2. Then:

- the mean of $Y_M(\zeta)$ is

$$\mathbb{E}\left[Y_M\right] = \mathbb{E}\left[\sum_{m=1}^{M} X_m(\zeta)\right] = \sum_{m=1}^{M} \mathbb{E}\left[X_m(\zeta)\right]$$

Central limit theorem

Consider the random variable $Y(\zeta)$ given by:

$$Y_M(\zeta) = \sum_{k=1}^{M} X_k(\zeta)$$

What is the distribution of $Y_M(\zeta)$ as $M \rightarrow \infty$?

Informally, the CLT is well known, and the answer is a Gaussian. Assume that the $X_M(\zeta)$'s are i. i. d., and the mean and variance of $X_m(\zeta)$ are finite and given by μ_X and σ_X^2. Then:

- the mean of $Y_M(\zeta)$ is

$$\mathbb{E}\left[Y_M\right] = \mathbb{E}\left[\sum_{m=1}^{M} X_m(\zeta)\right] = \sum_{m=1}^{M} \mathbb{E}\left[X_m(\zeta)\right]$$

$\mu_Y = M\mu_X$ What is μ_Y as $M \rightarrow \infty$?

Central limit theorem

Consider the random variable $Y(\zeta)$ given by:

$$Y_M(\zeta) = \sum_{k=1}^{M} X_k(\zeta)$$

What is the distribution of $Y_M(\zeta)$ as $M \rightarrow \infty$?

Informally, the CLT is well known, and the answer is a Gaussian. Assume that the $X_M(\zeta)$'s are i. i. d., and the mean and variance of $X_m(\zeta)$ are finite and given by μ_X and σ_X^2. Then:

- the variance of $Y_M(\zeta)$ is

$$\text{var}\left[Y_M\right] = \text{var}\left[\sum_{m=1}^{M} X_m(\zeta)\right] = \sum_{m=1}^{M} \text{var}\left[X_m(\zeta)\right]$$

$\sigma_Y^2 = M\sigma_X^2$ Similarly, what is σ_Y^2 as $M \rightarrow \infty$?

Central limit theorem

Theorem (Central limit theorem). Let $\{X_k(\zeta)\}_{k=1}^M$ be a collection of RVs that are independent and identically distributed for all $k = \{1, \dots, M\}$. Define the normalised random variable:

$$\hat{Y}_M(\zeta) = \frac{Y_M(\zeta) - \mu_{Y_M}}{\sigma_{Y_M}} \quad \text{where} \quad Y_M(\zeta) = \sum_{k=1}^{M} X_k(\zeta)$$

Then the distribution of $\hat{Y}_M(\zeta)$ approaches

$$\lim_{M\to\infty} f_{\hat{Y}_M}(y) = \mathcal{N}\left(y \,\middle|\, 0,\, 1\right)$$

♢

Central limit theorem

Theorem (Central limit theorem). PROOF. Since the $X_k(\zeta)$'s are i. i. d., then $\mu_{Y_M} = M\mu_X$ and $\sigma^2_{Y_M} = M\sigma^2_X$. Let

$$Z_k(\zeta) = \frac{X_k(\zeta) - \mu_X}{\sigma_X}$$

such that $\mu_{Z_k} = \mu_Z = 0$, $\sigma^2_{Z_k} = \sigma^2_Z = 1$ and:

$$\hat{Y}_M(\zeta) = \frac{1}{\sqrt{M}} \sum_{k=1}^{M} Z_k(\zeta)$$

□

Central limit theorem

Theorem (Central limit theorem). PROOF. Since the $X_k(\zeta)$'s are i. i. d., then $\mu_{Y_M} = M\mu_X$ and $\sigma^2_{Y_M} = M\sigma^2_X$. Let

$$Z_k(\zeta) = \frac{X_k(\zeta) - \mu_X}{\sigma_X}$$

such that $\mu_{Z_k} = \mu_Z = 0$, $\sigma^2_{Z_k} = \sigma^2_Z = 1$ and:

$$\hat{Y}_M(\zeta) = \frac{1}{\sqrt{M}} \sum_{k=1}^{M} Z_k(\zeta)$$

Noting that if $V(\zeta) = a\,U(\zeta)$ for some real-scalar a then

$$\Phi_V(\xi) = \mathbb{E}\left[e^{j\xi\, aU(\zeta)}\right] = \Phi_U(a\xi)$$

□

Central limit theorem

Theorem (Central limit theorem). PROOF. The normalised random variable can be written as:

$$\hat{Y}_M(\zeta) = \frac{1}{\sqrt{M}} \sum_{k=1}^{M} Z_k(\zeta)$$

Hence, the characteristic function for $\hat{Y}_M(\zeta)$ is given by:

$$\Phi_{\hat{Y}_M}(\xi) = \prod_{k=1}^{M} \Phi_{Z_k}\left(\frac{\xi}{\sqrt{M}}\right)$$

□

Central limit theorem

Theorem (Central limit theorem). PROOF. The normalised random variable can be written as:

$$\hat{Y}_M(\zeta) = \frac{1}{\sqrt{M}} \sum_{k=1}^{M} Z_k(\zeta)$$

Hence, the characteristic function for $\hat{Y}_M(\zeta)$ is given by:

$$\Phi_{\hat{Y}_M}(\xi) = \prod_{k=1}^{M} \Phi_{Z_k}\left(\frac{\xi}{\sqrt{M}}\right)$$

Since the $X_k(\zeta)$'s and therefore the $Z_k(\zeta)$'s are i. i. d., then $\Phi_{Z_k}(\xi) = \Phi_Z(\xi)$, or:

$$\Phi_{\hat{Y}_M}(\xi) = \Phi_Z^M\left(\frac{\xi}{\sqrt{M}}\right)$$

□

Central limit theorem

Theorem (Central limit theorem). PROOF. Since the $X_k(\zeta)$'s and therefore the $Z_k(\zeta)$'s are i. i. d., then $\Phi_{Z_k}(\xi) = \Phi_Z(\xi)$, or:

$$\Phi_{\hat{Y}_M}(\xi) = \Phi_Z^M\left(\frac{\xi}{\sqrt{M}}\right)$$

From the previous chapter on scalar random variables,

$$\Phi_Z(\xi) = \mathbb{E}\left[e^{j\xi\, Z(\zeta)}\right] = \sum_{n=0}^{\infty} \frac{(j\xi)^n}{n!}\mathbb{E}\left[Z^n(\zeta)\right]$$

□

Central limit theorem

Theorem (Central limit theorem). PROOF. From the previous chapter

$$\Phi_Z(\xi) = \mathbb{E}\left[e^{j\xi\, Z(\zeta)}\right] = \sum_{n=0}^{\infty} \frac{(j\xi)^n}{n!} \mathbb{E}\left[Z^n(\zeta)\right]$$

Therefore, the characteristic function for $\hat{Y}_M(\zeta)$ becomes:

$$\Phi_{\hat{Y}_M}(\xi) = \left\{\sum_{n=0}^{\infty} \frac{1}{n!}\left(\frac{j\xi}{\sqrt{M}}\right)^n \mathbb{E}\left[Z^n(\zeta)\right]\right\}^M$$

$$= \left\{1 + \frac{j\xi\mu_Z}{\sqrt{M}} - \frac{\xi^2\sigma_Z^2}{2M} + \mathcal{O}\left(\left\{\frac{\xi}{\sqrt{M}}\right\}^3\right)\right\}^M$$

□

Central limit theorem

Theorem (Central limit theorem). PROOF. Therefore, the characteristic function for $\hat{Y}_M(\zeta)$ becomes:

$$\Phi_{\hat{Y}_M}(\xi) = \left\{ \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{j\xi}{\sqrt{M}} \right)^n \mathbb{E}\left[Z^n(\zeta) \right] \right\}^M$$

$$= \left\{ 1 + \frac{j\xi\mu_Z}{\sqrt{M}} - \frac{\xi^2 \sigma_Z^2}{2M} + \mathcal{O}\left(\left\{ \frac{\xi}{\sqrt{M}} \right\}^3 \right) \right\}^M$$

Using the moments $\mu_Z = 0$ and $\sigma_Z^2 = 1$,

$$\Phi_{\hat{Y}_M}(\xi) = \left\{ 1 - \frac{\xi^2}{2M} + \mathcal{O}\left(\left\{ \frac{\xi}{\sqrt{M}} \right\}^3 \right) \right\}^M \rightarrow e^{-\frac{1}{2}\xi^2} \quad \text{as } M \rightarrow \infty$$

$\square$

Central limit theorem

Theorem (Central limit theorem). PROOF. Using the moments $\mu_Z = 0$ and $\sigma_Z^2 = 1$,

$$\Phi_{\hat{Y}_M}(\xi) = \left\{1 - \frac{\xi^2}{2M} + \mathcal{O}\left(\left\{\frac{\xi}{\sqrt{M}}\right\}^3\right)\right\}^M \rightarrow e^{-\frac{1}{2}\xi^2} \quad \text{as } M \rightarrow \infty$$

where the following limit is used:

$$\lim_{n\rightarrow\infty}\left(1 + \frac{x}{n}\right)^n = e^x$$

$\square$

Central limit theorem

Theorem (Central limit theorem). PROOF. Using the moments $\mu_Z = 0$ and $\sigma_Z^2 = 1$,

$$\Phi_{\hat{Y}_M}(\xi) = \left\{1 - \frac{\xi^2}{2M} + \mathcal{O}\left(\left\{\frac{\xi}{\sqrt{M}}\right\}^3\right)\right\}^M \to e^{-\frac{1}{2}\xi^2} \quad \text{as } M \to \infty$$

where the following limit is used:

$$\lim_{n\to\infty}\left(1 + \frac{x}{n}\right)^n = e^x \qquad \square$$

This last term is the characteristic function of the $\mathcal{N}\left(y \,\middle|\, 0,\, 1\right)$ distribution.

Central limit theorem

– End-of-Topic 35: **Central Limit Theorem** –

Any Questions?

Lecture Slideset 4
Estimation Theory

Introduction

- Thus far, have assumed that either the pdf or statistical values, such as mean, covariance, or higher order statistics, associated with a problem are fully known.

Introduction

- Thus far, have assumed that either the pdf or statistical values, such as mean, covariance, or higher order statistics, associated with a problem are fully known.

- In most practical applications, this is the exception rather than the rule.

Introduction

- Thus far, have assumed that either the pdf or statistical values, such as mean, covariance, or higher order statistics, associated with a problem are fully known.
- In most practical applications, this is the exception rather than the rule.
- The properties and parameters of random events must be obtained by collecting and analysing finite set of measurements.

Introduction

- Thus far, have assumed that either the pdf or statistical values, such as mean, covariance, or higher order statistics, associated with a problem are fully known.
- In most practical applications, this is the exception rather than the rule.
- The properties and parameters of random events must be obtained by collecting and analysing finite set of measurements.
- This handout will consider the problem of **Parameter Estimation**. This refers to the estimation of a parameter that is fixed, but is unknown.

A (Confusing) Note on Notation

Note that, unfortunately, from this point onwards, a slightly different notation for random quantities is used.

A (Confusing) Note on Notation

Note that, unfortunately, from this point onwards, a slightly different notation for random quantities is used.

- So far, particular **observations** of a random variable are written as lower-case letters, e.g. $x_n$or $x[n]$.

A (Confusing) Note on Notation

Note that, unfortunately, from this point onwards, a slightly different notation for random quantities is used.

- So far, particular **observations** of a random variable are written as lower-case letters, e.g. x_n or $x[n]$.
- Unfortunately, for convenience, lower-case letters are also used in some literature to refer to the **random variable** itself with the consequence that, in different contexts, $x[n]$ can refer to a particular observation, or a random value ($x[n] = X(\zeta)$).

A (Confusing) Note on Notation

Note that, unfortunately, from this point onwards, a slightly different notation for random quantities is used.

- So far, particular **observations** of a random variable are written as lower-case letters, e.g. $x_n$or $x[n]$.
- Unfortunately, for convenience, lower-case letters are also used in some literature to refer to the **random variable** itself with the consequence that, in different contexts, $x[n]$ can refer to a particular observation, or a random value ($x[n] = X(\zeta)$).
- The reason is due to the notation used to describe **random processes**, where the representation of a random process in the frequency domain is discussed, and upper-case letters are reserved to denote spectral representations.

A (Confusing) Note on Notation

Note that, unfortunately, from this point onwards, a slightly different notation for random quantities is used.

- So far, particular **observations** of a random variable are written as lower-case letters, e.g. $x_n$or $x[n]$.
- Unfortunately, for convenience, lower-case letters are also used in some literature to refer to the **random variable** itself with the consequence that, in different contexts, $x[n]$ can refer to a particular observation, or a random value ($x[n] = X(\zeta)$).

- The reason is due to the notation used to describe **random processes**, where the representation of a random process in the frequency domain is discussed, and upper-case letters are reserved to denote spectral representations.
- Moreover, lower-case letters for time-series helps with the clarity (where $x[n]$ is short-hand for $x[n, \zeta]$).

Examples of parameter estimation

Frequency Estimation Consider estimating the spectral content of a harmonic process, $x[n]$, consisting of a single-tone, given by

$$x[n] = A_0 \cos(\omega_0 n + \phi_0) + w[n]$$

where A_0, ϕ_0, and ω_0 are *unknown* constants, and where $w[n]$ is an additive white Gaussian noise (AWGN) process with zero-mean and variance σ^2. It is desired to estimate A_0, ϕ_0, and ω_0 from a realisation of the random process, giving rise to observations $x[n]$.

Examples of parameter estimation

Frequency Estimation Consider estimating the spectral content of a harmonic process, $x[n]$, consisting of a single-tone, given by

$$x[n] = A_0 \cos(\omega_0 n + \phi_0) + w[n]$$

where A_0, ϕ_0, and ω_0 are *unknown* constants, and where $w[n]$ is an AWGN process with zero-mean and variance σ^2. It is desired to estimate A_0, ϕ_0, and ω_0 from a realisation of the random process, giving rise to observations $x[n]$.

Sampling Distribution Parameters It is known that a set of observations, $\{x[n]\}_0^{N-1}$, are drawn from a sampling distribution with unknown parameters $\boldsymbol{\theta}$, such that:

$$x[n] \sim f_X\left(x \mid \boldsymbol{\theta}\right)$$

For example, if it is known that $x[n] \sim \mathcal{U}_{[a,\,b]}$, then it might be of interest to estimate the parameters a and b.

Examples of parameter estimation

Estimate of Moments It might be of interest to estimate the moments of a set of observations, $\{x[n]\}_0^{N-1}$, for example $\mu_X = \mathbb{E}\left[x[n]\right]$ and $\sigma_X^2 = \text{var}\left[x[n]\right]$.

Examples of parameter estimation

Estimate of Moments It might be of interest to estimate the moments of a set of observations, $\{x[n]\}_0^{N-1}$, for example $\mu_X = \mathbb{E}[x[n]]$ and $\sigma_X^2 = \text{var}[x[n]]$.

Constant value in noise An example which covers the various cases above is estimating a “direct current” (DC) constant in noise:

$$x[n] = A + w[n], \quad n \in \{0, \dots, N-1\}$$

Examples of parameter estimation

Estimate of Moments It might be of interest to estimate the moments of a set of observations, $\{x[n]\}_0^{N-1}$, for example $\mu_X = \mathbb{E}[x[n]]$ and $\sigma_X^2 = \text{var}[x[n]]$.

Constant value in noise An example which covers the various cases above is estimating a DC constant in noise:

$$x[n] = A + w[n], \quad n \in \{0, \ldots, N-1\}$$

- This list isn't exhaustiive, but gives an example of the type of **parameter estimation** problems that need to be addressed.

Properties of Estimators

Consider the set of N observations, $\mathcal{X} = \{x[n]\}_0^{N-1}$, from a *random experiment*; suppose they are used to estimate a parameter θ of the process using some function:

$$\hat{\theta} = \hat{\theta}\left[\mathcal{X}\right] = \hat{\theta}\left[\{x[n]\}_0^{N-1}\right]$$

Properties of Estimators

Consider the set of N observations, $\mathcal{X} = \{x[n]\}_0^{N-1}$, from a *random experiment*; suppose they are used to estimate a parameter θ of the process using some function:

$$\hat{\theta} = \hat{\theta}\left[\mathcal{X}\right] = \hat{\theta}\left[\{x[n]\}_0^{N-1}\right]$$

- The function $\hat{\theta}\left[\mathcal{X}\right]$ is known as an **estimator** whereas the value taken by the estimator, using a particular set of observations, is called a **point-estimate**.

Properties of Estimators

Consider the set of N observations, $\mathcal{X} = \{x[n]\}_0^{N-1}$, from a *random experiment*; suppose they are used to estimate a parameter θ of the process using some function:

$$\hat{\theta} = \hat{\theta}\left[\mathcal{X}\right] = \hat{\theta}\left[\{x[n]\}_0^{N-1}\right]$$

- The function $\hat{\theta}\left[\mathcal{X}\right]$ is known as an **estimator** whereas the value taken by the estimator, using a particular set of observations, is called a **point-estimate**.

- An aim is to design an estimator, $\hat{\theta}$, that should be as close to the true value of the parameter, θ, as possible.

Properties of Estimators

Since $\hat{\theta}$ is a function of a number of realisations of a random experiment, it is itself a RV, and thus has a mean and variance.

Properties of Estimators

Since $\hat{\theta}$ is a function of a number of realisations of a random experiment, it is itself a RV, and thus has a mean and variance.

- As an example of an estimator, consider estimating the mean μ_X of a random variate, $X(\zeta)$, from N observations $\mathcal{X} = \{x[n]\}_0^{N-1}$. The most natural estimator is a simple arithmetic average of these observations, the **sample mean**:

$$\hat{\mu}_X = \hat{\theta}[\mathcal{X}] = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$$

Properties of Estimators

Since $\hat{\theta}$ is a function of a number of realisations of a random experiment, it is itself a RV, and thus has a mean and variance.

- As an example of an estimator, consider estimating the mean μ_X of a random variate, $X(\zeta)$, from N observations $\mathcal{X} = \{x[n]\}_0^{N-1}$. The most natural estimator is a simple arithmetic average of these observations, the **sample mean**:

$$\hat{\mu}_X = \hat{\theta}[\mathcal{X}] = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$$

- To demonstrate that these estimates are RVs, consider repeating the procedure for calculating the sample mean from a large number of difference sets of realisations.
  - Then a large number of estimates of μ_X, denoted by the set $\{\hat{\mu}_X\}$, is obtained, and these can be used to generate a histogram showing the distribution of the estimates.

Properties of Estimators

Example (Numerical Example). Suppose that $N = 1000$ observations are generated from a Gaussian density with mean $\mu = 5$ and variance $\sigma^2 = 1$. Use MATLAB and a Monte Carlo experiment to find the distribution of the sample mean.

SOLUTION. One realisation would generate $N = 1000$ data points generated from $x[n] \sim \mathcal{N}\left(\mu = 5,\, \sigma^2 = 1\right)$ using:

```
mu = 5; sigma = 1; N = 1000;
x = mu + sigma * randn(N, 1);
muEst = sum(x)/N
```

Properties of Estimators

Example (Numerical Example). Suppose that $N = 1000$ observations are generated from a Gaussian density with mean $\mu = 5$ and variance $\sigma^2 = 1$. Use MATLAB and a Monte Carlo experiment to find the distribution of the sample mean.

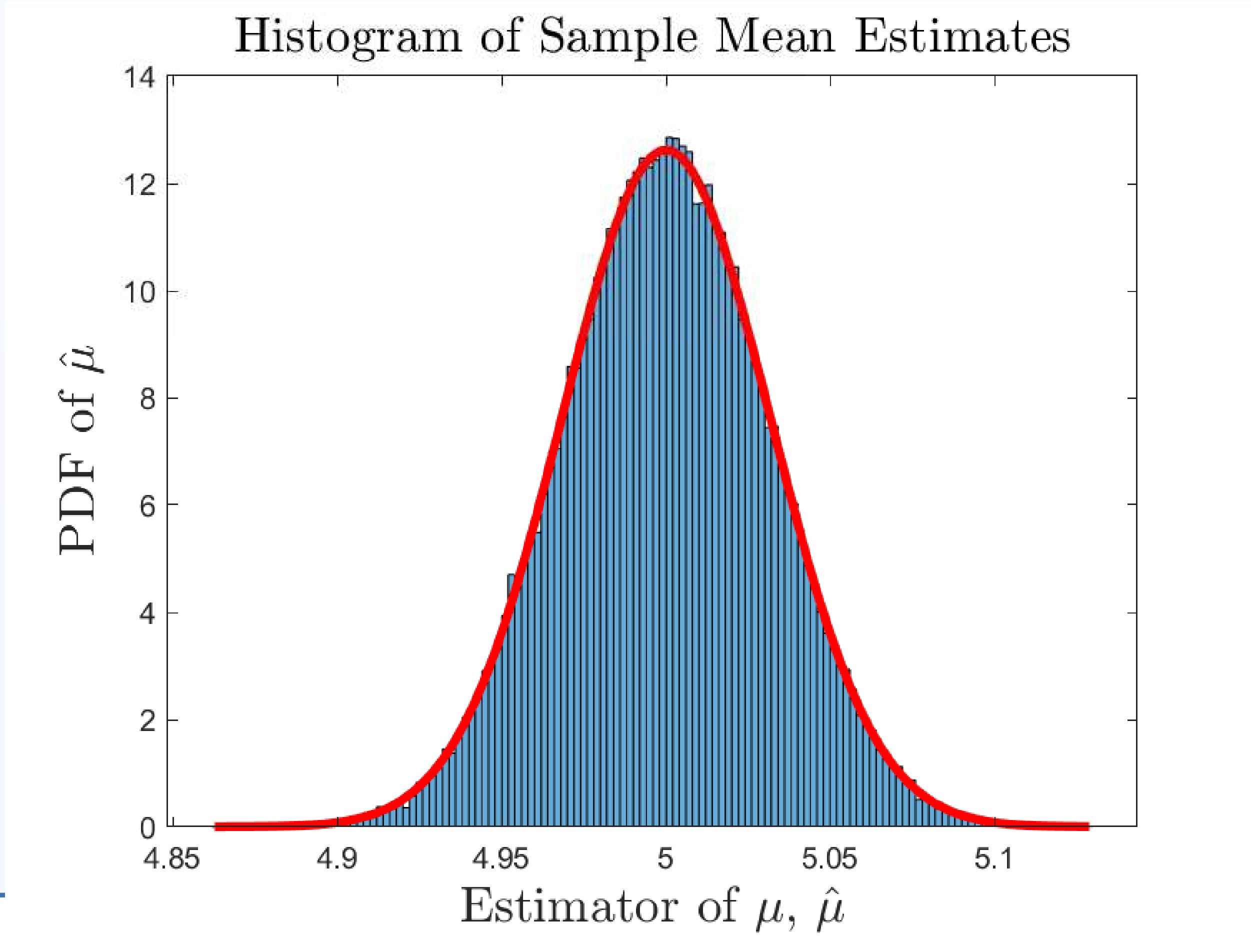
SOLUTION. This can be repeated $K = 100000$ times to produce a Monte Carlo estimate. This can be achieved with the following code: □

```
N = 1000; K = 100000;
mu = 5; sigma = 1;
muEst = zeros(1, K);
for k = 1 : K
x = mu + sigma * randn(N, 1);
muEst(k) = sum(x) / N;
end
mean(muEst)
```

Properties of Estimators

Example (Numerical Example). Use MATLAB and a Monte Carlo experiment to find the distribution of the sample mean.

SOLUTION. The results of this Monte Carlo experiment are:

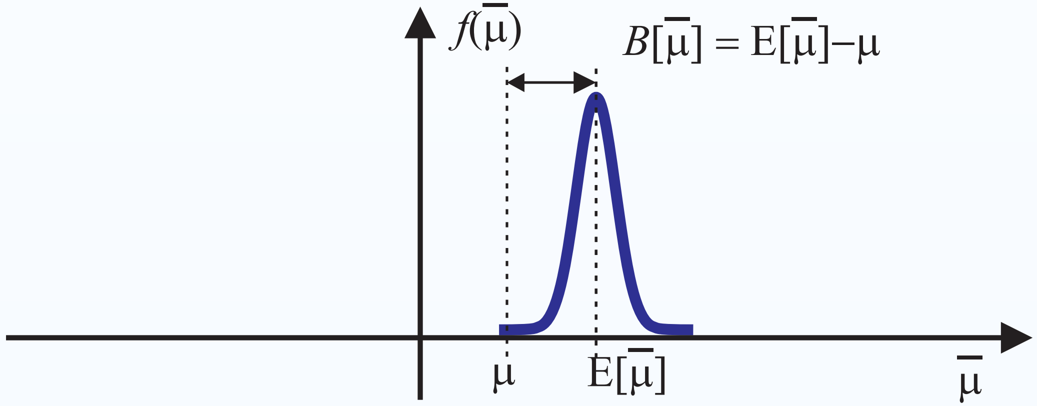


Properties of Estimators

– End-of-Topic 36: **Introduction to Estimation Theory and the Definition of an Estimator** –

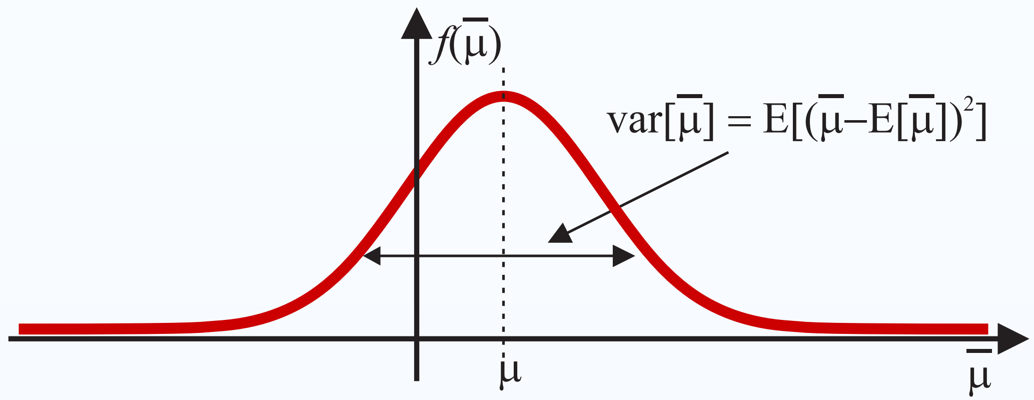
Any Questions?

What makes a good estimator?



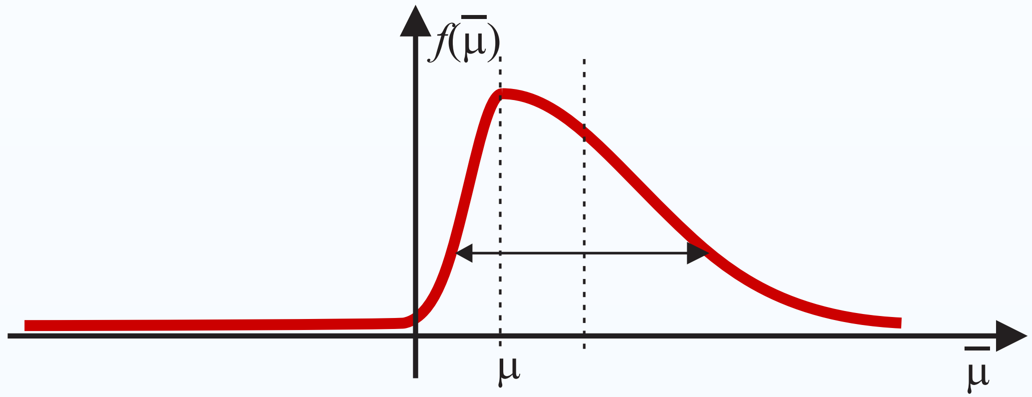
Here, the pdf of the estimated value, $\bar{\mu}$, is biased away from the true value, μ. However, the spread of the estimated value around the true value is small.

What makes a good estimator?



Here, the pdf of the estimated value, $\bar{\mu}$, is centered on the true value, μ. However, the spread of the estimated value around the true value is very large.

What makes a good estimator?



It is important to note that higher-order statistics can also play a part in quantifying the performance of an estimator, although that won't be considered further here.

Bias of estimator

The **bias** of an estimator $\hat{\theta}$ of a parameter θ is defined as:

$$B(\hat{\theta}) \triangleq \mathbb{E}\left[\hat{\theta}\right] - \theta$$

Bias of estimator

The **bias** of an estimator $\hat{\theta}$ of a parameter θ is defined as:

$$B(\hat{\theta}) \triangleq \mathbb{E}\left[\hat{\theta}\right] - \theta$$

If θ is large, then a small deviation would give what would appear to be a large bias. Therefore, the **normalised bias** is therefore often used instead:

$$\epsilon_b(\hat{\theta}) \triangleq \frac{B(\hat{\theta})}{\theta} = \frac{\mathbb{E}\left[\hat{\theta}\right]}{\theta} - 1, \quad \theta \neq 0$$

Bias of estimator

The **bias** of an estimator $\hat{\theta}$ of a parameter θ is defined as:

$$B(\hat{\theta}) \triangleq \mathbb{E}\left[\hat{\theta}\right] - \theta$$

If θ is large, then a small deviation would give what would appear to be a large bias. Therefore, the **normalised bias** is therefore often used instead:

$$\epsilon_b(\hat{\theta}) \triangleq \frac{B(\hat{\theta})}{\theta} = \frac{\mathbb{E}\left[\hat{\theta}\right]}{\theta} - 1, \quad \theta \neq 0$$

Example (Biasness of sample mean estimator). Is the sample mean, $\hat{\mu}_x = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$ biased?

SOLUTION. No, since

$$\mathbb{E}\left[\hat{\mu}_x\right] = \mathbb{E}\left[\frac{1}{N}\sum_{n=0}^{N-1} x[n]\right] = \frac{1}{N}\sum_{n=0}^{N-1} \mathbb{E}\left[x[n]\right] = \frac{N\mu_X}{N} = \mu_X.$$

Variance of estimator

The **variance** of the estimator $\hat{\theta}$ is defined by:

$$\operatorname{var}\left[\hat{\theta}\right] = \sigma_{\hat{\theta}}^2 \triangleq \mathbb{E}\left[\left|\hat{\theta} - \mathbb{E}\left[\hat{\theta}\right]\right|^2\right]$$

However, a minimum variance criterion is not always compatible with the minimum bias requirement; reducing the variance may result in an increase in bias.

Variance of estimator

The **variance** of the estimator $\hat{\theta}$ is defined by:

$$\operatorname{var}\left[\hat{\theta}\right] = \sigma^2_{\hat{\theta}} \triangleq \mathbb{E}\left[\left|\hat{\theta} - \mathbb{E}\left[\hat{\theta}\right]\right|^2\right]$$

However, a minimum variance criterion is not always compatible with the minimum bias requirement; reducing the variance may result in an increase in bias.

Therefore, a compromise or balance between these two conflicting criteria is required, and this is provided by the mean-squared error (MSE) measure described in the next topic.

Variance of estimator

The **variance** of the estimator $\hat{\theta}$ is defined by:

$$\text{var}\left[\hat{\theta}\right] = \sigma_{\hat{\theta}}^2 \triangleq \mathbb{E}\left[\left|\hat{\theta} - \mathbb{E}\left[\hat{\theta}\right]\right|^2\right]$$

However, a minimum variance criterion is not always compatible with the minimum bias requirement; reducing the variance may result in an increase in bias.

Therefore, a compromise or balance between these two conflicting criteria is required, and this is provided by the mean-squared error (MSE) measure described in the next topic.

The **normalised standard deviation** is defined by:

$$\epsilon_r \triangleq \frac{\sigma_{\hat{\theta}}}{\theta}, \quad \theta \neq 0$$

Variance of estimator

Example (Variance of Sample Mean). Calculate the variance of the sample mean, assuming the observations are independent.

Variance of estimator

Example (Variance of Sample Mean). Calculate the variance of the sample mean, assuming the observations are independent.

SOLUTION. Noting $\{x[n]\}_{n=0}^{N-1}$ are i. i. d. with variance σ_x^2, then there are two approaches to calculating the variance.

Variance of estimator

Example (Variance of Sample Mean). Calculate the variance of the sample mean, assuming the observations are independent.

SOLUTION. Noting $\{x[n]\}_{n=0}^{N-1}$ are i. i. d. with variance σ_x^2, then there are two approaches to calculating the variance.

- The first is to use the result that:

$$\text{var}\left[\sum_{n=0}^{N-1} c_n \, X_n(\zeta)\right] = \sum_{n=0}^{N-1} c_n^2 \, \text{var}\left[X_n(\zeta)\right]$$

□

Variance of estimator

Example (Variance of Sample Mean). Calculate the variance of the sample mean, assuming the observations are independent.

SOLUTION. Noting $\{x[n]\}_{n=0}^{N-1}$ are i. i. d. with variance σ_x^2, then there are two approaches to calculating the variance.

- The first is to use the result that:

$$\text{var}\left[\sum_{n=0}^{N-1} c_n\, X_n(\zeta)\right] = \sum_{n=0}^{N-1} c_n^2\ \text{var}\left[X_n(\zeta)\right]$$

Therefore,

$$\text{var}\left[\hat{\mu}_x\right] = \text{var}\left[\frac{1}{N}\sum_{n=0}^{N-1} x[n]\right] = \frac{1}{N^2}\sum_{n=0}^{N-1} \text{var}\left[x[n]\right] = \frac{\sigma_x^2}{N}$$

□

Variance of estimator

Example (Variance of Sample Mean). Calculate the variance of the sample mean, assuming the observations are independent.

SOLUTION. The second approach uses the result that

$$\mathbb{E}\left[x[n]\,x[m]\right] = \sigma_x^2\,\delta(n-m) + \mu_x^2.$$

□

Variance of estimator

Example (Variance of Sample Mean). Calculate the variance of the sample mean, assuming the observations are independent.

SOLUTION. ● The second approach uses the result that $\mathbb{E}\left[x[n]\, x[m]\right] = \sigma_x^2\, \delta(n-m) + \mu_x^2$.

● The sample mean estimator is unbiased, and therefore writing $\theta = \mu_x$, then $\mathbb{E}\left[\hat{\mu}_x\right] = \mu_x$. Therefore:

$$\operatorname{var}\left[\hat{\mu}_x\right] = \mathbb{E}\left[\left|\left\{\frac{1}{N}\sum_{n=0}^{N-1} x[n]\right\} - \mu_x\right|^2\right]$$

□

Variance of estimator

Example (Variance of Sample Mean). Calculate the variance of the sample mean, assuming the observations are independent.

SOLUTION.
- The second approach uses the result that $\mathbb{E}\left[x[n]\,x[m]\right] = \sigma_x^2\,\delta(n-m) + \mu_x^2$.
- The sample mean estimator is unbiased, and therefore writing $\theta = \mu_x$, then $\mathbb{E}\left[\hat{\mu}_x\right] = \mu_x$. Therefore:

$$\operatorname{var}\left[\hat{\mu}_x\right] = \mathbb{E}\left[\left|\left\{\frac{1}{N}\sum_{n=0}^{N-1} x[n]\right\} - \mu_x\right|^2\right]$$

$$= \mathbb{E}\left[\frac{1}{N^2}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} x[n]\,x[m] - 2\frac{\mu_x}{N}\sum_{n=0}^{N-1} x[n] + \mu_x^2\right]$$

□

Variance of estimator

Example (Variance of Sample Mean). Calculate the variance of the sample mean, assuming the observations are independent.

SOLUTION.
- The second approach uses the result that $\mathbb{E}\left[x[n]\,x[m]\right] = \sigma_x^2\,\delta(n-m) + \mu_x^2$.
- The sample mean estimator is unbiased, and therefore writing $\theta = \mu_x$, then $\mathbb{E}\left[\hat{\mu}_x\right] = \mu_x$. Therefore:

$$\operatorname{var}\left[\hat{\mu}_x\right] = \mathbb{E}\left[\left|\left\{\frac{1}{N}\sum_{n=0}^{N-1} x[n]\right\} - \mu_x\right|^2\right]$$

$$= \mathbb{E}\left[\frac{1}{N^2}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} x[n]\,x[m] - 2\frac{\mu_x}{N}\sum_{n=0}^{N-1} x[n] + \mu_x^2\right]$$

$$= \frac{1}{N^2}\left\{N\left[\sigma_x^2 + N\mu_x^2\right] - 2N^2\mu_x^2 + N^2\mu_x^2\right\} = \frac{\sigma_x^2}{N} \qquad \square$$

Variance of estimator

– End-of-Topic 37: **What makes a good estimator? Introduction to bias and variance** –

Any Questions?

Mean square error

Minimising estimator variance can increase bias. A compromise criterion is the MSE of the estimator, which is given by:

$$\text{MSE}(\hat{\theta}) = \mathbb{E}\left[\left|\hat{\theta} - \theta\right|^2\right] = \sigma_{\hat{\theta}}^2 + |B(\hat{\theta})|^2$$

Mean square error

Minimising estimator variance can increase bias. A compromise criterion is the MSE of the estimator, which is given by:

$$\text{MSE}(\hat{\theta}) = \mathbb{E}\left[\left|\hat{\theta} - \theta\right|^2\right] = \sigma_{\hat{\theta}}^2 + |B(\hat{\theta})|^2$$

- The estimator $\hat{\theta}_{\text{MSE}} = \hat{\theta}_{\text{MSE}}[\mathcal{X}]$ which minimises $\text{MSE}(\hat{\theta})$ is known as the minimum mean-square error:

$$\hat{\theta}_{MSE} = \arg_{\hat{\theta}} \min \text{MSE}(\hat{\theta})$$

Mean square error

Minimising estimator variance can increase bias. A compromise criterion is the MSE of the estimator, which is given by:

$$\text{MSE}(\hat{\theta}) = \mathbb{E}\left[\left|\hat{\theta} - \theta\right|^2\right] = \sigma^2_{\hat{\theta}} + |B(\hat{\theta})|^2$$

- The estimator $\hat{\theta}_{\text{MSE}} = \hat{\theta}_{\text{MSE}}[\mathcal{X}]$ which minimises $\text{MSE}(\hat{\theta})$ is known as the minimum mean-square error:

$$\hat{\theta}_{MSE} = \arg_{\hat{\theta}} \min \text{MSE}(\hat{\theta})$$

- This measures the average mean squared deviation of the estimator from its true value.
- Unfortunately, adoption of this natural criterion leads to unrealisable estimators; ones which cannot be written solely as a function of the data.

Mean square error

Example ([Kay:1993, Example 2.1, Pages 16 and 19]). Consider the observations

$$x[n] = A + w[n]\,, \quad n \in \{0, \dots, N-1\}$$

where A is the parameter to be estimated, and $w[n]$ is white Gaussian noise (WGN) with variance σ^2. A reasonable estimator for the average value of $x[n]$, A, is:

$$\hat{A}_a = a\,\frac{1}{N}\sum_{n=0}^{N-1} x[n]$$

- If $a = 1$, then this is just the sample mean.
- Find the optimal (modified) estimator $\hat{A}_a$ by finding the value of a that minimises the MSE. ⋈

Mean square error

Example ([Kay:1993, Example 2.1, Pages 16 and 19]). Consider

$$x[n] = A + w[n]\,, \quad n \in \{0, \dots, N-1\}$$

A reasonable estimator for A, is:

$$\hat{A}_a = a\,\frac{1}{N}\sum_{n=0}^{N-1} x[n]$$

- Find the optimal $\hat{A}_a$ by finding a that minimises the MSE.

SOLUTION. Due to the linearity properties of the expectation operator, then it can be seen, as in the previous example, that:

$$\mathbb{E}\left[\hat{A}_a\right] = \mathbb{E}\left[a\frac{1}{N}\sum_{n=0}^{N-1} x[n]\right] = aA$$

Mean square error

Example ([Kay:1993, Example 2.1, Pages 16 and 19]). Find the optimal $\hat{A}_a$ by finding a that minimises the MSE.

SOLUTION. Therefore, this is a **biased estimate** with bias $B\left(\hat{A}_a\right) = A(a-1)$. As in the previous example, then:

$$\operatorname{var}\left[\hat{A}_a\right] = \operatorname{var}\left[a\frac{1}{N}\sum_{n=0}^{N-1} x[n]\right]$$

$$= \frac{a^2}{N^2}\sum_{n=0}^{N-1}\operatorname{var}\left[x[n]\right] = \frac{a^2\sigma^2}{N}$$

□

Mean square error

Example ([Kay:1993, Example 2.1, Pages 16 and 19]). Find the optimal $\hat{A}_a$ by finding a that minimises the MSE.

SOLUTION. Therefore, this is a **biased estimate** with bias $B\left(\hat{A}_a\right) = A(a-1)$. As in the previous example, then:

$$\mathrm{var}\left[\hat{A}_a\right] = \mathrm{var}\left[a\frac{1}{N}\sum_{n=0}^{N-1} x[n]\right]$$

$$= \frac{a^2}{N^2}\sum_{n=0}^{N-1}\mathrm{var}\left[x[n]\right] = \frac{a^2\sigma^2}{N}$$

Hence, the MSE is given by:

$$\mathrm{MSE}(\hat{A}_a) = \mathrm{var}\left[\hat{A}_a\right] + |B(\hat{A}_a)|^2 = \frac{a^2\sigma^2}{N} + (a-1)^2A^2$$

Mean square error

Example ([Kay:1993, Example 2.1, Pages 16 and 19]). Find the optimal $\hat{A}_a$ by finding a that minimises the MSE.

SOLUTION. Hence, the MSE is given by:

$$\text{MSE}(\hat{A}_a) = \text{var}\left[\hat{A}_a\right] + |B(\hat{A}_a)|^2 = \frac{a^2\sigma^2}{N} + (a-1)^2 A^2$$

In order to find the minimum mean-square error (MMSE), then differentiate this and set to zero:

$$\frac{d\text{MSE}(\hat{A}_a)}{da} = \frac{2a\sigma^2}{N} + 2(a-1)A^2$$

□

Mean square error

Example ([Kay:1993, Example 2.1, Pages 16 and 19]). Find the optimal $\hat{A}_a$ by finding a that minimises the MSE.

SOLUTION. Hence, the MSE is given by:

$$\mathrm{MSE}(\hat{A}_a) = \mathrm{var}\left[\hat{A}_a\right] + |B(\hat{A}_a)|^2 = \frac{a^2\sigma^2}{N} + (a-1)^2 A^2$$

In order to find the minimum mean-square error (MMSE), then differentiate this and set to zero:

$$\frac{d\mathrm{MSE}(\hat{A}_a)}{da} = \frac{2a\sigma^2}{N} + 2(a-1)A^2$$

which is equal to zero when

$$a_{\mathrm{opt}} = \frac{A^2}{A^2 + \frac{\sigma^2}{N}}$$

Mean square error

Example ([Kay:1993, Example 2.1, Pages 16 and 19]). Find the optimal $\hat{A}_a$ by finding a that minimises the MSE.

SOLUTION. In order to find the minimum mean-square error (MMSE), then differentiate this and set to zero:

$$a_{\text{opt}} = \frac{A^2}{A^2 + \frac{\sigma^2}{N}}$$

- Thus, the optimal value of a depends upon the unknown parameter A.
- The estimator is therefore not realisable, and this is since the bias term is a function of A.
- Any criterion which depends on the bias of the estimator will, generally, lead to an unrealisable estimator. On occasion realisable MMSE estimators can be found. □

Mean square error

Example ([Kay:1993, Example 2.1, Pages 16 and 19]). Find the optimal $\hat{A}_a$ by finding a that minimises the MSE.

SOLUTION. Despite the unrealisable estimator, the result can still be informative. First, note that:

$$a_{\text{opt}} = \frac{1}{1 + \frac{1}{N}\left(\frac{\sigma^2}{A^2}\right)} = \frac{1}{1 + \frac{1}{N\,\text{SNR}}}$$

where the signal-to-noise ratio (SNR) is: $\text{SNR} = \frac{A^2}{\sigma^2}$.

- It is apparent that when N and the SNR are low, some value less than $a = 1$ may be appropriate. □

Mean square error

Example ([Kay:1993, Example 2.1, Pages 16 and 19]). Find the optimal $\hat{A}_a$ by finding a that minimises the MSE.

SOLUTION. Despite the unrealisable estimator, the result can still be informative. First, note that:

$$a_{\text{opt}} = \frac{1}{1 + \frac{1}{N}\left(\frac{\sigma^2}{A^2}\right)} = \frac{1}{1 + \frac{1}{N\,\text{SNR}}}$$

where the SNR is: $\text{SNR} = \frac{A^2}{\sigma^2}$.

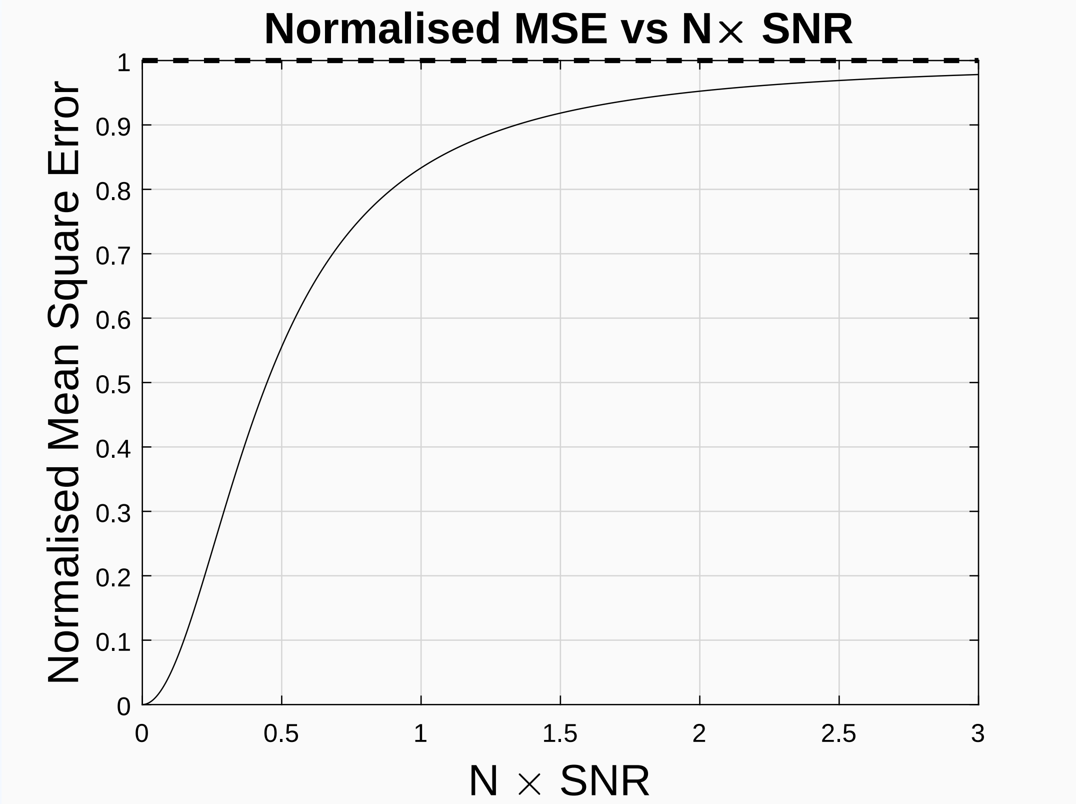
- The minimum MSE can be calculated as:

$$\text{MSE}\left(a_{\text{opt}}\right) = \frac{\sigma^2}{N}\left(\frac{1}{1 + \frac{1}{N\,\text{SNR}}}\right)$$

Mean square error

Example ([Kay:1993, Example 2.1, Pages 16 and 19]). Find the optimal $\hat{A}_a$ by finding a that minimises the MSE.

SOLUTION. Despite the unrealisable estimator, the result can still be informative.

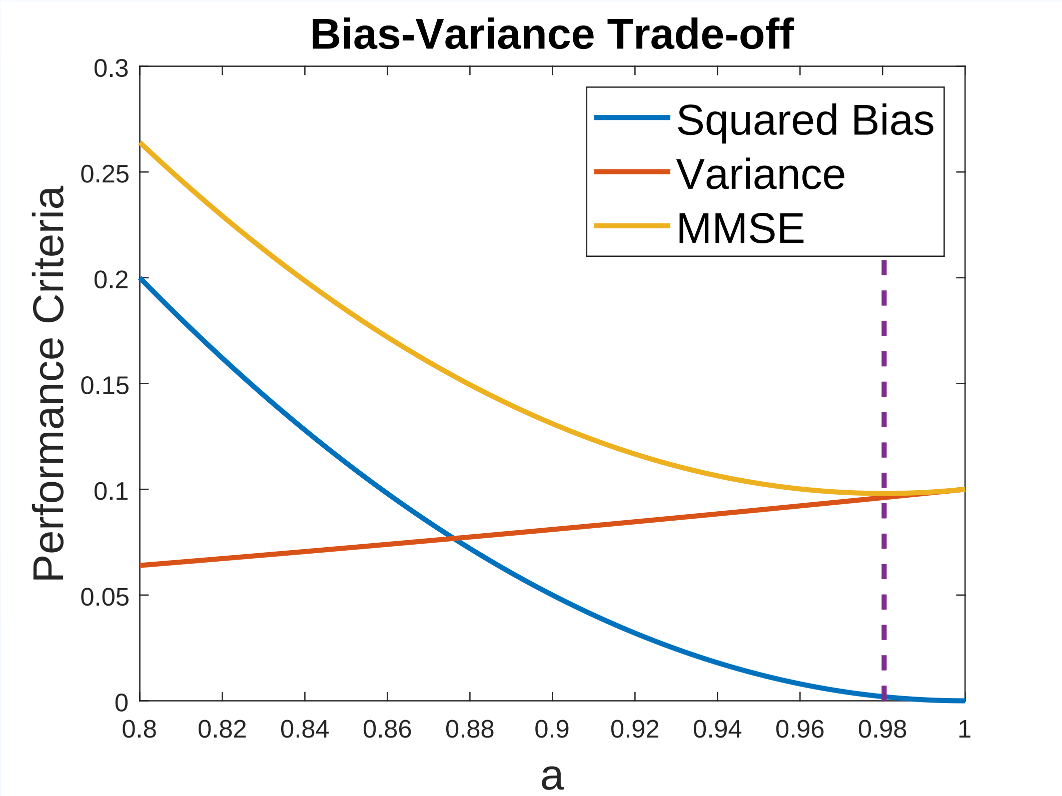


□

Mean square error

Example ([Kay:1993, Example 2.1, Pages 16 and 19]). Find the optimal $\hat{A}_a$ by finding a that minimises the MSE.

SOLUTION. Moreover, by plotting the bias, variance, and MSE, we can see how the bias-variance trade-off occurs.



□

Mean square error

– End-of-Topic 38: **Mean Square Error and MSE Estimators** –

Any Questions?

Consistency of an Estimator

If the MSE of the estimator,

$$\text{MSE}(\hat{\theta}) = \mathbb{E}\left[|\hat{\theta} - \theta|^2\right] = \sigma^2_{\hat{\theta}} + |B(\hat{\theta})|^2$$

approaches zero as the sample size N becomes large, then both the bias and the variance tends toward zero.

Consistency of an Estimator

If the MSE of the estimator,

$$\text{MSE}(\hat{\theta}) = \mathbb{E}\left[|\hat{\theta} - \theta|^2\right] = \sigma^2_{\hat{\theta}} + |B(\hat{\theta})|^2$$

approaches zero as the sample size N becomes large, then both the bias and the variance tends toward zero.

- Thus, the sampling distribution tends to concentrate around θ, and as $N \to \infty$, it will become an impulse at θ.

Consistency of an Estimator

If the MSE of the estimator,

$$\mathrm{MSE}(\hat{\theta}) = \mathbb{E}\left[|\hat{\theta} - \theta|^2\right] = \sigma_{\hat{\theta}}^2 + |B(\hat{\theta})|^2$$

approaches zero as the sample size N becomes large, then both the bias and the variance tends toward zero.

- Thus, the sampling distribution tends to concentrate around θ, and as $N \to \infty$, it will become an impulse at θ.
- This is a very important and desirable property, and such an estimator is called a **consistent estimator**.

Consistency of an Estimator

If the MSE of the estimator,

$$\text{MSE}(\hat{\theta}) = \mathbb{E}\left[|\hat{\theta} - \theta|^2\right] = \sigma_{\hat{\theta}}^2 + |B(\hat{\theta})|^2$$

approaches zero as the sample size N becomes large, then both the bias and the variance tends toward zero.

- Thus, the sampling distribution tends to concentrate around θ, and as $N \to \infty$, it will become an impulse at θ.
- This is a very important and desirable property, and such an estimator is called a **consistent estimator**.

Definition (Efficiency of an estimator). An estimate is said to be **efficient** w. r. t. another estimate if it has a lower variance. Thus, if $\hat{\theta}_N$ is an estimator that depends on N observations and is both **unbiased** and **efficient** with respect to $\hat{\theta}_{N-1}$ for all N, then $\hat{\theta}_N$ is a **consistent** estimate.

Consistency of an Estimator

– End-of-Topic 39: **Consistency of Estimator** –

Any Questions?

Cramer-Rao Lower Bound

- In the previous Topic, the performance of a given estimator has been considered; what is the bias, and what is the variance?
- The MSE criterion gives a possible design method for finding the structural form of an optimal estimator, but isn't always realisable.

Cramer-Rao Lower Bound

- In the previous Topic, the performance of a given estimator has been considered; what is the bias, and what is the variance?
- The MSE criterion gives a possible design method for finding the structural form of an optimal estimator, but isn't always realisable.
- This leads to the general question of whether there is a particular methodology for designing an estimator for a given probabilistic problem.

Cramer-Rao Lower Bound

- In the previous Topic, the performance of a given estimator has been considered; what is the bias, and what is the variance?
- The MSE criterion gives a possible design method for finding the structural form of an optimal estimator, but isn't always realisable.
- This leads to the general question of whether there is a particular methodology for designing an estimator for a given probabilistic problem.

- If the MSE can be minimised when the bias is zero, then clearly the variance is also minimised. Such estimators are called MVUEs.

Cramer-Rao Lower Bound

- In the previous Topic, the performance of a given estimator has been considered; what is the bias, and what is the variance?
- The MSE criterion gives a possible design method for finding the structural form of an optimal estimator, but isn't always realisable.
- This leads to the general question of whether there is a particular methodology for designing an estimator for a given probabilistic problem.

- If the MSE can be minimised when the bias is zero, then clearly the variance is also minimised. Such estimators are called MVUEs.
- MVUE possess the important property that they attain a minimum bound on the variance of the estimator, called the Cramér-Rao lower-bound (CRLB).

Cramer-Rao Lower Bound

Theorem (CRLB - real scalar parameter). If $\mathbf{X}(\zeta) = [x[0], \cdots, x[N-1]]^T$ and $f_{\mathbf{X}}(\mathbf{x} \mid \theta)$ is the joint density of $\mathbf{X}(\zeta)$ which depends on the fixed but unknown parameter θ, the variance of $\hat{\theta}$ is bounded by:

$$\operatorname{var}\left[\hat{\theta}\right] \geq \frac{1}{\mathbb{E}\left[\left(\frac{\partial \ln f_{\mathbf{X}}(\mathbf{x} \mid \theta)}{\partial \theta}\right)^2\right]}$$

$\diamondsuit$

Cramer-Rao Lower Bound

Theorem (CRLB - real scalar parameter). If $\mathbf{X}(\zeta) = [x[0], \cdots, x[N-1]]^T$ and $f_{\mathbf{X}}(\mathbf{x} \mid \theta)$ is the joint density of $\mathbf{X}(\zeta)$ which depends on the fixed but unknown parameter θ, the variance of $\hat{\theta}$ is bounded by:

$$\text{var}\left[\hat{\theta}\right] \geq \frac{1}{\mathbb{E}\left[\left(\frac{\partial \ln f_{\mathbf{X}}(\mathbf{x} \mid \theta)}{\partial \theta}\right)^2\right]}$$

Alternatively, it may also be expressed as:

$$\text{var}\left[\hat{\theta}\right] \geq -\frac{1}{\mathbb{E}\left[\frac{\partial^2 \ln f_{\mathbf{X}}(\mathbf{x} \mid \theta)}{\partial \theta^2}\right]}$$

$\diamond$

Cramer-Rao Lower Bound

Theorem (CRLB - real scalar parameter). If $\mathbf{X}(\zeta) = [x[0], \cdots, x[N-1]]^T$ and $f_{\mathbf{X}}(\mathbf{x} \,|\, \theta)$ is the joint density of $\mathbf{X}(\zeta)$ which depends on the fixed but unknown parameter θ, the variance of $\hat{\theta}$ is bounded by:

$$\mathrm{var}\left[\hat{\theta}\right] \geq \frac{1}{\mathbb{E}\left[\left(\frac{\partial \ln f_{\mathbf{X}}(\mathbf{x} \,|\, \theta)}{\partial \theta}\right)^2\right]}$$

Alternatively, it may also be expressed as:

$$\mathrm{var}\left[\hat{\theta}\right] \geq -\frac{1}{\mathbb{E}\left[\frac{\partial^2 \ln f_{\mathbf{X}}(\mathbf{x} \,|\, \theta)}{\partial \theta^2}\right]} \qquad \diamondsuit$$

The function $\ln f_{\mathbf{X}}(\mathbf{x} \,|\, \theta)$ is called the **log-likelihood** of θ.

Cramer-Rao Lower Bound

Theorem (CRLB - real scalar parameter). If $\mathbf{X}(\zeta) = [x[0], \cdots, x[N-1]]^T$ and $f_{\mathbf{X}}(\mathbf{x} \mid \theta)$ is the joint density of $\mathbf{X}(\zeta)$ which depends on the fixed but unknown parameter θ, the variance of $\hat{\theta}$ is bounded by:

$$\operatorname{var}\left[\hat{\theta}\right] \geq \frac{1}{\mathbb{E}\left[\left(\frac{\partial \ln f_{\mathbf{X}}(\mathbf{x} \mid \theta)}{\partial \theta}\right)^2\right]}$$

Alternatively, it may also be expressed as:

$$\operatorname{var}\left[\hat{\theta}\right] \geq -\frac{1}{\mathbb{E}\left[\frac{\partial^2 \ln f_{\mathbf{X}}(\mathbf{x} \mid \theta)}{\partial \theta^2}\right]}$$

Furthermore, an unbiased estimator may be found that attains the bound for all θ if, and only if, (iff)

$$\frac{\partial \ln f_{\mathbf{X}}(\mathbf{x} \mid \theta)}{\partial \theta} = I(\theta)\left(\hat{\theta} - \theta\right)$$

◇

Cramer-Rao Lower Bound

Example ([Kay:1993, Example 3.3, Page 31]). Consider again:

$$x[n] = A + w[n]\,, \quad n \in \{0, \dots, N-1\}$$ ⋈

where A is the parameter to be estimated, and $w[n]$ is WGN. Determine the CRLB for an estimator, $\hat{A}$, of the parameter A.

Cramer-Rao Lower Bound

Example ([Kay:1993, Example 3.3, Page 31]). Consider again:

$$x[n] = A + w[n]\,, \quad n \in \{0, \dots, N-1\}$$

where A is the parameter to be estimated, and $w[n]$ is WGN. Determine the CRLB for an estimator, $\hat{A}$, of the parameter A.

SOLUTION. Since the transformation between $w[n]$ and $x[n]$ is linear, with a multiplication factor of 1, the *likelihood function*is:

$$f_{\mathbf{X}}(\mathbf{x} \mid A) = \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x[n]-A)^2\right]$$

$$= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right]$$

□

Cramer-Rao Lower Bound

Example ([Kay:1993, Example 3.3, Page 31]). Consider again:

$$x[n] = A + w[n]\,, \quad n \in \{0, \dots, N-1\}$$

where A is the parameter to be estimated, and $w[n]$ is WGN. Determine the CRLB for an estimator, $\hat{A}$, of the parameter A.

SOLUTION. Taking the first derivative of the **log-likelihood**:

$$\frac{\partial \ln f_{\mathbf{x}}\left(\mathbf{x} \mid A\right)}{\partial A} = \frac{\partial}{\partial A}\left[-\frac{N}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-A\right)^2\right]$$

□

Cramer-Rao Lower Bound

Example ([Kay:1993, Example 3.3, Page 31]). Consider again:

$$x[n] = A + w[n]\,, \quad n \in \{0, \dots, N-1\}$$

where A is the parameter to be estimated, and $w[n]$ is WGN. Determine the CRLB for an estimator, $\hat{A}$, of the parameter A.

SOLUTION. Taking the first derivative of the **log-likelihood**:

$$\frac{\partial \ln f_{\mathbf{X}}(\mathbf{x} \mid A)}{\partial A} = \frac{\partial}{\partial A}\left[-\frac{N}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right]$$

$$= \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x[n]-A) = \frac{N}{\sigma^2}\left(\left\{\frac{1}{N}\sum_{n=0}^{N-1}x[n]\right\} - A\right)$$

□

Cramer-Rao Lower Bound

Example ([Kay:1993, Example 3.3, Page 31]). Consider again:

$$x[n] = A + w[n], \quad n \in \{0, \dots, N-1\}$$

where A is the parameter to be estimated, and $w[n]$ is WGN. Determine the CRLB for an estimator, $\hat{A}$, of the parameter A.

SOLUTION. Taking the first derivative of the **log-likelihood**:

$$\frac{\partial \ln f_{\mathbf{X}}(\mathbf{x} \mid A)}{\partial A} = \frac{\partial}{\partial A}\left[-\frac{N}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right]$$

$$= \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x[n]-A) = \frac{N}{\sigma^2}\left(\left\{\frac{1}{N}\sum_{n=0}^{N-1}x[n]\right\} - A\right)$$

$$= \frac{N}{\sigma^2}\left(\hat{\mu}_X - A\right)$$

□

Cramer-Rao Lower Bound

Example ([Kay:1993, Example 3.3, Page 31]). Consider again:

$$x[n] = A + w[n]\,, \quad n \in \{0, \dots, N-1\}$$

where A is the parameter to be estimated, and $w[n]$ is WGN. Determine the CRLB for an estimator, $\hat{A}$, of the parameter A.

SOLUTION. Differentiating again, then:

$$\frac{\partial^2 \ln f_{\mathbf{X}}\left(\mathbf{x} \mid A\right)}{\partial A^2} = -\frac{N}{\sigma^2}$$

□

Cramer-Rao Lower Bound

Example ([Kay:1993, Example 3.3, Page 31]). Consider again:

$$x[n] = A + w[n], \quad n \in \{0, \dots, N-1\}$$

where A is the parameter to be estimated, and $w[n]$ is WGN. Determine the CRLB for an estimator, $\hat{A}$, of the parameter A.

SOLUTION. Differentiating again, then:

$$\frac{\partial^2 \ln f_{\mathbf{X}}(\mathbf{x} \mid A)}{\partial A^2} = -\frac{N}{\sigma^2}$$

and noting that this is constant, then the CRLB is:

$$\text{var}\left[\hat{A}\right] \geq \frac{\sigma^2}{N}$$

□

Cramer-Rao Lower Bound

Example ([Kay:1993, Example 3.3, Page 31]). Consider again:

$$x[n] = A + w[n]\,, \quad n \in \{0, \dots, N-1\}$$

where A is the parameter to be estimated, and $w[n]$ is WGN. Determine the CRLB for an estimator, $\hat{A}$, of the parameter A.

SOLUTION. It is noted the first derivative of the log-likelihood is in the form:

$$\frac{\partial \ln f_{\mathbf{X}}\left(\mathbf{x} \mid \theta\right)}{\partial \theta} = I(\theta)\left(\hat{\theta} - \theta\right) = \frac{N}{\sigma^2}\left(\left\{\frac{1}{N}\sum_{n=0}^{N-1} x[n]\right\} - A\right)$$

□

Cramer-Rao Lower Bound

Example ([Kay:1993, Example 3.3, Page 31]). Consider again:

$$x[n] = A + w[n]\,, \quad n \in \{0, \dots, N-1\}$$

where A is the parameter to be estimated, and $w[n]$ is WGN. Determine the CRLB for an estimator, $\hat{A}$, of the parameter A.

SOLUTION. It is noted the first derivative of the log-likelihood is in the form:

$$\frac{\partial \ln f_{\mathbf{X}}\left(\mathbf{x} \mid \theta\right)}{\partial \theta} = I(\theta)\left(\hat{\theta} - \theta\right) = \frac{N}{\sigma^2}\left(\left\{\frac{1}{N}\sum_{n=0}^{N-1} x[n]\right\} - A\right) \quad \square$$

then it is clear that the sample mean attains the bound, such that $\hat{A} = \mu_X$, and must therefore be the MVUE. Hence, the minimum variance will also be given by $\operatorname{var}\left[\hat{A}\right] = \frac{\sigma^2}{N}$.

Cramer-Rao Lower Bound

– End-of-Topic 40: **Introduction to the CRLB and how to identify MVUE that satisfy the bound** –

Any Questions?

Maximum Likelihood Estimation

The joint density of the RVs $\mathbf{X}(\zeta) = \{x[n, \zeta]\}_0^{N-1}$, which depends on fixed but unknown parameter $\boldsymbol{\theta}$, is $f_{\mathbf{X}}(\mathbf{x} \mid \boldsymbol{\theta})$.

Maximum Likelihood Estimation

The joint density of the RVs $\mathbf{X}(\zeta) = \{x[n, \zeta]\}_0^{N-1}$, which depends on fixed but unknown parameter $\boldsymbol{\theta}$, is $f_{\mathbf{X}}\left(\mathbf{x} \mid \boldsymbol{\theta}\right)$.

- This same quantity, viewed as a function of the parameter $\boldsymbol{\theta}$ when a particular set of observations, $\hat{\mathbf{x}}$ is given, is known as the **likelihood function**.

Maximum Likelihood Estimation

The joint density of the RVs $\mathbf{X}(\zeta) = \{x[n,\zeta]\}_0^{N-1}$, which depends on fixed but unknown parameter $\boldsymbol{\theta}$, is $f_{\mathbf{X}}\left(\mathbf{x} \mid \boldsymbol{\theta}\right)$.

- This same quantity, viewed as a function of the parameter $\boldsymbol{\theta}$ when a particular set of observations, $\hat{\mathbf{x}}$ is given, is known as the **likelihood function**.

- The **maximum-likelihood estimate (MLE)** of the parameter $\boldsymbol{\theta}$, denoted by $\hat{\boldsymbol{\theta}}_{ml}$, is defined as that value of $\boldsymbol{\theta}$ that maximises $f_{\mathbf{X}}\left(\hat{\mathbf{x}} \mid \boldsymbol{\theta}\right)$.

Maximum Likelihood Estimation

The joint density of the RVs $\mathbf{X}(\zeta) = \{x[n,\zeta]\}_0^{N-1}$, which depends on fixed but unknown parameter $\boldsymbol{\theta}$, is $f_{\mathbf{X}}(\mathbf{x} \mid \boldsymbol{\theta})$.

- This same quantity, viewed as a function of the parameter $\boldsymbol{\theta}$ when a particular set of observations, $\hat{\mathbf{x}}$ is given, is known as the **likelihood function**.
- The **maximum-likelihood estimate (MLE)** of the parameter $\boldsymbol{\theta}$, denoted by $\hat{\boldsymbol{\theta}}_{ml}$, is defined as that value of $\boldsymbol{\theta}$ that maximises $f_{\mathbf{X}}(\hat{\mathbf{x}} \mid \boldsymbol{\theta})$.
- The MLE for $\boldsymbol{\theta}$ is defined by:

$$\hat{\boldsymbol{\theta}}_{ml}(\mathbf{x}) = \arg_{\boldsymbol{\theta}} \max f_{\mathbf{X}}(\mathbf{x} \mid \boldsymbol{\theta})$$

- Note that since $\hat{\boldsymbol{\theta}}_{ml}(\mathbf{x})$ depends on the random observation vector $\mathbf{x}$, and so is *itself a RV*.

Maximum Likelihood Estimation

Assuming a differentiable likelihood function, and that $\boldsymbol{\theta} \in \mathbb{R}^P$, the MLE is found from

$$\begin{bmatrix} \frac{\partial f_{\mathbf{X}}(\mathbf{x} \mid \boldsymbol{\theta})}{\partial \theta_1} \\ \vdots \\ \frac{\partial f_{\mathbf{X}}(\mathbf{x} \mid \boldsymbol{\theta})}{\partial \theta_P} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

or, more simply,

$$\nabla_{\boldsymbol{\theta}} f_{\mathbf{X}}(\mathbf{x} \mid \boldsymbol{\theta}) \triangleq \frac{\partial f_{\mathbf{X}}(\mathbf{x} \mid \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}_{P \times 1}$$

where $\mathbf{0}_{P\times 1}$ denotes the $P \times 1$ vector of zero elements. If multiple solutions to this exist, then the one that maximises the likelihood function is the MLE.

Properties of the MLE

1. The MLE satisfies

$$\nabla_{\boldsymbol{\theta}} f_{\mathbf{X}}\left(\mathbf{x} \mid \boldsymbol{\theta}\right)\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{ml}} = \mathbf{0}_{P\times 1}$$

$$\nabla_{\boldsymbol{\theta}} \ln f_{\mathbf{X}}\left(\mathbf{x} \mid \boldsymbol{\theta}\right)\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{ml}} = \mathbf{0}_{P\times 1}$$

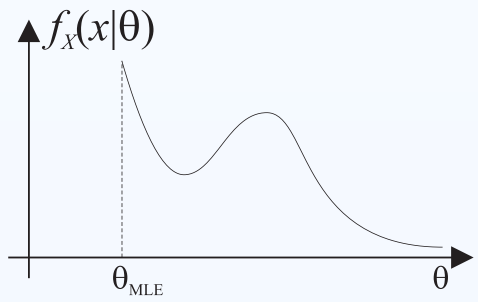
Properties of the MLE

1. The MLE satisfies

$$\nabla_{\boldsymbol{\theta}} f_{\mathbf{X}}\left(\mathbf{x} \mid \boldsymbol{\theta}\right)\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{ml}} = \mathbf{0}_{P\times 1}$$

$$\nabla_{\boldsymbol{\theta}} \ln f_{\mathbf{X}}\left(\mathbf{x} \mid \boldsymbol{\theta}\right)\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{ml}} = \mathbf{0}_{P\times 1}$$

2. If an MVUE exists and the MLE does not occur at a boundary, then the MLE *is* the MVUE.



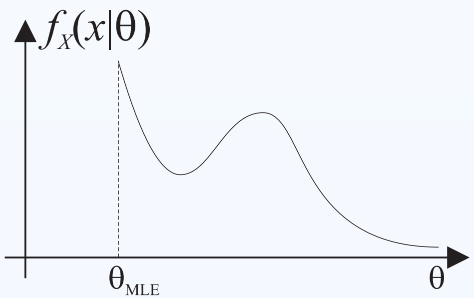
A single parameter MLE that occurs at a boundary

Properties of the MLE

1. The MLE satisfies

$$\nabla_{\boldsymbol{\theta}} f_{\mathbf{X}}\left(\mathbf{x} \mid \boldsymbol{\theta}\right)\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{ml}} = \mathbf{0}_{P\times 1}$$
$$\nabla_{\boldsymbol{\theta}} \ln f_{\mathbf{X}}\left(\mathbf{x} \mid \boldsymbol{\theta}\right)\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{ml}} = \mathbf{0}_{P\times 1}$$

2. If an MVUE exists and the MLE does not occur at a boundary, then the MLE *is* the MVUE.



A single parameter MLE that occurs at a boundary

3. MLE is asymptotically distributed according to a Gaussian:

$$\hat{\boldsymbol{\theta}}_{ml} \sim \mathcal{N}\left(\boldsymbol{\theta}, \mathbf{J}^{-1}(\boldsymbol{\theta})\right)$$

DC Level in white Gaussian noise

Example ([Therrien:1991, Example 6.1, Page 282]). A constant but unknown signal is observed in additive WGN. That is,

$$x[n] = A + w[n] \quad \text{where} \quad w[n] \sim \mathcal{N}\left(0,\, \sigma_w^2\right) \qquad \bowtie$$

for $n \in \mathcal{N} = \{0, \dots, N-1\}$. Calculate the MLE of A.

DC Level in white Gaussian noise

Example ([Therrien:1991, Example 6.1, Page 282]). A constant but unknown signal is observed in additive WGN. That is,

$$x[n] = A + w[n] \quad \text{where} \quad w[n] \sim \mathcal{N}\left(0,\, \sigma_w^2\right)$$

for $n \in \mathcal{N} = \{0, \dots, N-1\}$. Calculate the MLE of A.

SOLUTION. Since this is a memoryless system, and $w[n]$ are i. i. d., then so is $x[n]$, and therefore:

$$\ln f_{\mathbf{x}}\left(\mathbf{x} \mid A\right) = -\frac{N}{2}\ln(2\pi\sigma_w^2) - \frac{\sum_{n\in\mathcal{N}}\left(x[n]-A\right)^2}{2\sigma_w^2}$$

□

DC Level in white Gaussian noise

Example ([Therrien:1991, Example 6.1, Page 282]). A constant but unknown signal is observed in additive WGN. That is,

$$x[n] = A + w[n] \quad \text{where} \quad w[n] \sim \mathcal{N}\left(0,\, \sigma_w^2\right)$$

for $n \in \mathcal{N} = \{0, \dots, N-1\}$. Calculate the MLE of A.

SOLUTION. Since this is a memoryless system, and $w[n]$ are i. i. d., then so is $x[n]$, and therefore:

$$\ln f_{\mathbf{X}}\left(\mathbf{x} \mid A\right) = -\frac{N}{2}\ln(2\pi\sigma_w^2) - \frac{\sum_{n\in\mathcal{N}}\left(x[n]-A\right)^2}{2\sigma_w^2}$$

Differentiating this expression w. r. t. A

$$\frac{\partial \ln f_{\mathbf{X}}\left(\mathbf{x} \mid A\right)}{\partial A} = \frac{\sum_{n\in\mathcal{N}}\left(x[n]-A\right)}{\sigma_w^2}$$

□

DC Level in white Gaussian noise

Example ([Therrien:1991, Example 6.1, Page 282]). A constant but unknown signal is observed in additive WGN. That is,

$$x[n] = A + w[n] \quad \text{where} \quad w[n] \sim \mathcal{N}\left(0,\, \sigma_w^2\right)$$

for $n \in \mathcal{N} = \{0, \dots, N-1\}$. Calculate the MLE of A.

SOLUTION. Since this is a memoryless system, and $w[n]$ are i. i. d., then so is $x[n]$, and therefore:

$$\ln f_{\mathbf{X}}\left(\mathbf{x} \mid A\right) = -\frac{N}{2}\ln(2\pi\sigma_w^2) - \frac{\sum_{n\in\mathcal{N}}\left(x[n]-A\right)^2}{2\sigma_w^2}$$

Differentiating this expression w. r. t. A and setting to zero:

$$\hat{A}_{ml} = \frac{1}{N}\sum_{n\in\mathcal{N}} x[n]$$

DC Level in white Gaussian noise

Example ([Therrien:1991, Example 6.1, Page 282]). A constant but unknown signal is observed in additive WGN. That is,

$$x[n] = A + w[n] \quad \text{where} \quad w[n] \sim \mathcal{N}\left(0,\, \sigma_w^2\right)$$

for $n \in \mathcal{N} = \{0, \dots, N-1\}$. Calculate the MLE of A.

SOLUTION. Differentiating this expression w. r. t. A and setting to zero:

$$\hat{A}_{ml} = \frac{1}{N} \sum_{n \in \mathcal{N}} x[n]$$

- This is the **sample mean**, and it has already been seen that this is an efficient estimator. Hence, the MLE is efficient.
- This result is true in general; if an **efficient estimator** exists, the *maximum likelihood procedure* will produce it. □

DC Level in white Gaussian noise

Example ([Therrien:1991, Example 6.1, Page 282]). A constant but unknown signal is observed in additive WGN. That is,

$$x[n] = A + w[n] \quad \text{where} \quad w[n] \sim \mathcal{N}\left(0,\, \sigma_w^2\right)$$

for $n \in \mathcal{N} = \{0, \dots, N-1\}$. Calculate the MLE of A.

SOLUTION. To complete the solution, check this does, in fact, correspond to a maximum rather than a minimum or other stationary point.

DC Level in white Gaussian noise

Example ([Therrien:1991, Example 6.1, Page 282]). A constant but unknown signal is observed in additive WGN. That is,

$$x[n] = A + w[n] \quad \text{where} \quad w[n] \sim \mathcal{N}\left(0,\, \sigma_w^2\right)$$

for $n \in \mathcal{N} = \{0, \dots, N-1\}$. Calculate the MLE of A.

SOLUTION. To complete the solution, check this does, in fact, correspond to a maximum rather than a minimum or other stationary point.

- This can be verified by differentiating for a second time:

$$\frac{\partial^2 \ln f_{\mathbf{X}}\left(\mathbf{x} \mid A\right)}{\partial A^2} = \frac{\sum_{n \in \mathcal{N}} (-1)}{\sigma_w^2} = \frac{-N}{\sigma_w^2} < 0$$

- which is always negative and therefore corresponds to a minimum. □

MLE for Transformed Parameter

Theorem (Invariance Property of the MLE). The MLE of the parameter $\boldsymbol{\alpha} = \mathbf{g}(\boldsymbol{\theta})$, where $\mathbf{g}$ is an r-dimensional function of the $P \times 1$ parameter $\boldsymbol{\theta}$, and the pdf, $f_{\mathbf{X}}(\mathbf{x} \mid \boldsymbol{\theta})$ is parameterised by $\boldsymbol{\theta}$, is given by

$$\hat{\boldsymbol{\alpha}}_{ml} = \mathbf{g}(\hat{\boldsymbol{\theta}}_{ml}) \qquad \diamondsuit$$

where $\hat{\boldsymbol{\theta}}_{ml}$ is the MLE of $\boldsymbol{\theta}$.

MLE for Transformed Parameter

Theorem (Invariance Property of the MLE). The MLE of the parameter $\boldsymbol{\alpha} = \mathbf{g}(\boldsymbol{\theta})$, where $\mathbf{g}$ is an r-dimensional function of the $P \times 1$ parameter $\boldsymbol{\theta}$, and the pdf, $f_{\mathbf{X}}(\mathbf{x} \mid \boldsymbol{\theta})$ is parameterised by $\boldsymbol{\theta}$, is given by

$$\hat{\boldsymbol{\alpha}}_{ml} = \mathbf{g}(\hat{\boldsymbol{\theta}}_{ml})$$

where $\hat{\boldsymbol{\theta}}_{ml}$ is the MLE of $\boldsymbol{\theta}$.

- The MLE of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}_{ml}$, is obtained by maximising $f_{\mathbf{X}}(\mathbf{x} \mid \boldsymbol{\theta})$. $\diamondsuit$

MLE for Transformed Parameter

Theorem (Invariance Property of the MLE). The MLE of the parameter $\boldsymbol{\alpha} = \mathrm{g}(\boldsymbol{\theta})$, where g is an r-dimensional function of the $P \times 1$ parameter $\boldsymbol{\theta}$, and the pdf, $f_{\mathbf{X}}\left(\mathbf{x} \mid \boldsymbol{\theta}\right)$ is parameterised by $\boldsymbol{\theta}$, is given by

$$\hat{\boldsymbol{\alpha}}_{ml} = \mathbf{g}(\hat{\boldsymbol{\theta}}_{ml})$$

where $\hat{\boldsymbol{\theta}}_{ml}$ is the MLE of $\boldsymbol{\theta}$.

- The MLE of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}_{ml}$, is obtained by maximising $f_{\mathbf{X}}\left(\mathbf{x} \mid \boldsymbol{\theta}\right)$.
- If the function g is not an invertible function, then $\hat{\boldsymbol{\alpha}}$ maximises the modified likelihood function $\bar{p}_T\left(\mathbf{x} \mid \boldsymbol{\alpha}\right)$ defined as:

$$\bar{p}_T\left(\mathbf{x} \mid \boldsymbol{\alpha}\right) = \max_{\boldsymbol{\theta}:\boldsymbol{\alpha}=\mathbf{g}(\boldsymbol{\theta})} f_{\mathbf{X}}\left(\mathbf{x} \mid \boldsymbol{\theta}\right) \qquad \diamond$$

MLE for Transformed Parameter

– End-of-Topic 41: **Introduction to MLE** –

Any Questions?

Least Squares

The estimators discussed so far have attempted to find an optimal or nearly optimal (for large data records) estimator for example, the MVUE.

Least Squares

The estimators discussed so far have attempted to find an optimal or nearly optimal (for large data records) estimator for example, the MVUE.

- For some techniques, this means that the pdf of the data must be known somehow.
- An alternate philosophy is a class of estimators that in general have no optimality properties associated with them, but make *good sense* for many problems: the **principle of least squares**.

Least Squares

The estimators discussed so far have attempted to find an optimal or nearly optimal (for large data records) estimator for example, the MVUE.

- For some techniques, this means that the pdf of the data must be known somehow.
- An alternate philosophy is a class of estimators that in general have no optimality properties associated with them, but make *good sense* for many problems: the **principle of least squares**.
- A salient feature of the method is that *no probabilistic assumptions* are made about the data; only a *signal model* is assumed.

Least Squares

The estimators discussed so far have attempted to find an optimal or nearly optimal (for large data records) estimator for example, the MVUE.

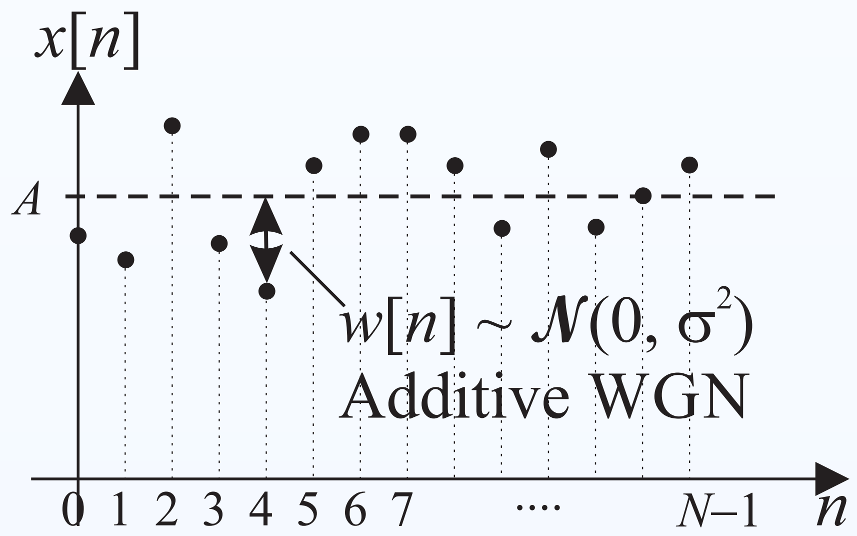
- For some techniques, this means that the pdf of the data must be known somehow.
- An alternate philosophy is a class of estimators that in general have no optimality properties associated with them, but make *good sense* for many problems: the **principle of least squares**.
- A salient feature of the method is that *no probabilistic assumptions* are made about the data; only a *signal model* is assumed.
- As will be seen, it turns out that the LSE can be calculated when just the first and second moments are known, and through the solution of *linear* equations.

The Least Squares Approach

In the least-squares (LS) approach, it is sought to minimise the squared difference between the given, or observed, data $x[n]$ and the assumed, or hidden, signal or noiseless data.

The Least Squares Approach

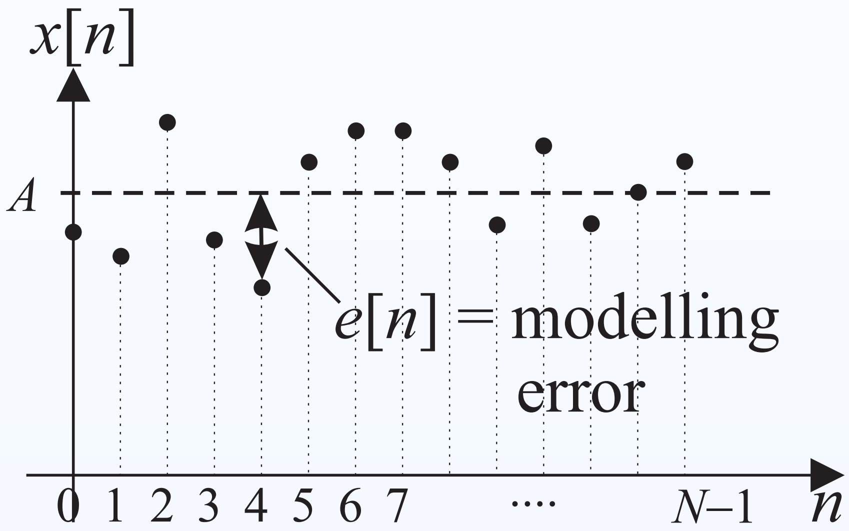
In the LS approach, it is sought to minimise the squared difference between the given, or observed, data $x[n]$ and the assumed, or hidden, signal or noiseless data.



- In the MLE method, the observed data $x[n] \equiv x[n, \zeta]$ is considered to be a random variable consisting of a known signal model, denoted $s[n; \boldsymbol{\theta}]$, where $\boldsymbol{\theta}$ is a set of unknown model parameters, plus a noise term, $w[n, \zeta]$, with a given pdf.

The Least Squares Approach

In the LS approach, it is sought to minimise the squared difference between the given, or observed, data $x[n]$ and the assumed, or hidden, signal or noiseless data.



- In contrast to the MLE method, the least squares method considers $x[n]$ to be the sum of a known signal model, $s[n; \boldsymbol{\theta}]$, plus an error term $e[n]$.
- This error term really consists of two components: the modelling error, and an observation error.

The Least Squares Approach

In the LS approach, it is sought to minimise the squared difference between the given, or observed, data $x[n]$ and the assumed, or hidden, signal or noiseless data.

- Here it is assumed that the signal is generated by some model which, in turn, depends on some unknown parameter $\boldsymbol{\theta}$.

The Least Squares Approach

In the LS approach, it is sought to minimise the squared difference between the given, or observed, data $x[n]$ and the assumed, or hidden, signal or noiseless data.

- Here it is assumed that the signal is generated by some model which, in turn, depends on some unknown parameter $\boldsymbol{\theta}$.

- Now, one approach to finding the estimator is to minimise the sum of the absolute errors:

$$\hat{\boldsymbol{\theta}}_{L_1} = \arg_{\boldsymbol{\theta}} \min J_1(\boldsymbol{\theta}) \quad \text{where} \quad J_1(\boldsymbol{\theta}) = \sum_{n=0}^{N-1} |x[n] - s[n, \boldsymbol{\theta}]|$$

- However, in practice, while this is a good optimisation problem to solve, this is a difficult calculation in many cases.

The Least Squares Approach

In the LS approach, it is sought to minimise the squared difference between the given, or observed, data $x[n]$ and the assumed, or hidden, signal or noiseless data.

- Here it is assumed that the signal is generated by some model which, in turn, depends on some unknown parameter $\boldsymbol{\theta}$.
- The LSE of $\boldsymbol{\theta}$ chooses the value that makes $s[n]$ closest to data $x[n]$, and this *closeness* is measured by the LS error criterion:

$$J(\boldsymbol{\theta}) = \sum_{n=0}^{N-1} (x[n] - s[n])^2$$

The Least Squares Approach

In the LS approach, it is sought to minimise the squared difference between the given, or observed, data $x[n]$ and the assumed, or hidden, signal or noiseless data.

- Here it is assumed that the signal is generated by some model which, in turn, depends on some unknown parameter $\boldsymbol{\theta}$.
- The LSE of $\boldsymbol{\theta}$ chooses the value that makes $s[n]$ closest to data $x[n]$, and this *closeness* is measured by the LS error criterion:

$$J(\boldsymbol{\theta}) = \sum_{n=0}^{N-1} \left(x[n] - s[n]\right)^2$$

- The LSE is given by:

$$\hat{\boldsymbol{\theta}}_{LSE} = \arg_{\boldsymbol{\theta}} \min J(\boldsymbol{\theta})$$

DC Level

Example (Sample mean revisited). It is assumed that an observed signal, $x[n]$, is a perturbed version of an unknown signal, $s[n]$, which is modelled as $s[n] = A$, for $n \in \mathcal{N} = \{0, \dots, N-1\}$. Calculate the LSE of the unknown signal A.

DC Level

Example (Sample mean revisited). It is assumed that an observed signal, $x[n]$, is a perturbed version of an unknown signal, $s[n]$, which is modelled as $s[n] = A$, for $n \in \mathcal{N} = \{0, \dots, N-1\}$. Calculate the LSE of the unknown signal A.

SOLUTION. According to the LS approach, then:

$$\hat{A}_{LSE} = \arg_A \min J(A) \quad \text{where} \quad J(A) = \sum_{n=0}^{N-1} (x[n] - A)^2$$

Differentiating w. r. t. A and setting the result to zero produces

$$\hat{A}_{LSE} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] \qquad \square$$

which is the sample mean estimator.

DC Level

Example (Sample mean revisited). It is assumed that an observed signal, $x[n]$, is a perturbed version of an unknown signal, $s[n]$, which is modelled as $s[n] = A$, for $n \in \mathcal{N} = \{0, \dots, N-1\}$. Calculate the LSE of the unknown signal A.

SOLUTION. According to the LS approach, then:

$$\hat{A}_{LSE} = \arg_A \min J(A) \quad \text{where} \quad J(A) = \sum_{n=0}^{N-1} (x[n] - A)^2$$

Differentiating w. r. t. A and setting the result to zero produces

$$\hat{A}_{LSE} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] \qquad \square$$

Differentiating for a second time shows this indeed minimises the squared error.

Nonlinear Least Squares

Example (Sinusoidal Frequency Estimation). Again, it is assumed that an observed signal, $x[n]$, is a perturbed version of an unknown signal, $s[n]$, which is modelled as

$$s[n] = \cos 2\pi f_0 n \qquad \bowtie$$

in which the frequency f_0 is to be estimated.

Nonlinear Least Squares

Example (Sinusoidal Frequency Estimation). Again, it is assumed that an observed signal, $x[n]$, is a perturbed version of an unknown signal, $s[n]$, which is modelled as

$$s[n] = \cos 2\pi f_0 n$$

in which the frequency f_0 is to be estimated.

- The LSE can be found by minimising:

$$J(f_0) = \sum_{n=0}^{N-1} (x[n] - \cos 2\pi f_0 n)^2$$

$\bowtie$

Nonlinear Least Squares

Example (Sinusoidal Frequency Estimation). Again, it is assumed that an observed signal, $x[n]$, is a perturbed version of an unknown signal, $s[n]$, which is modelled as

$$s[n] = \cos 2\pi f_0 n$$

in which the frequency f_0 is to be estimated.

- The LSE can be found by minimising:

$$J(f_0) = \sum_{n=0}^{N-1} (x[n] - \cos 2\pi f_0 n)^2$$

- The LS error function is highly nonlinear in the parameter f_0.
- The minimisation cannot be done in closed form. $\bowtie$

Nonlinear Least Squares

Example (Sinusoidal Frequency Estimation). Again, it is assumed that an observed signal, $x[n]$, is a perturbed version of an unknown signal, $s[n]$, which is modelled as

$$s[n] = \cos 2\pi f_0 n$$

in which the frequency f_0 is to be estimated.

- The LSE can be found by minimising:

$$J(f_0) = \sum_{n=0}^{N-1} (x[n] - \cos 2\pi f_0 n)^2$$

- The LS error function is highly nonlinear in the parameter f_0.
- The minimisation cannot be done in closed form.
- A signal model that is *linear in the unknown parameter* is said to generate a **linear least squares** problem. $\bowtie$

Nonlinear Least Squares

Example (Sinusoidal Frequency Estimation). Again, it is assumed that an observed signal, $x[n]$, is a perturbed version of an unknown signal, $s[n]$, which is modelled as

$$s[n] = \cos 2\pi f_0 n$$

in which the frequency f_0 is to be estimated.

- The LSE can be found by minimising:

$$J(f_0) = \sum_{n=0}^{N-1} \left(x[n] - \cos 2\pi f_0 n\right)^2$$

- The LS error function is highly nonlinear in the parameter f_0.
- The minimisation cannot be done in closed form.
- **Nonlinear least squares** problems are solved via grid searches or iterative minimisation methods. ⋈

Nonlinear Least Squares

– End-of-Topic 42: **Introduction to Least Squares Estimation** –

Any Questions?

Linear Least Squares

Assume that an observed signal, $\{x[n]\}_0^{N-1}$, is a perturbed version of an unknown signal, $\{s[n]\}_0^{N-1}$, where each of these processes can be written by the random vectors:

$$\mathbf{s} = \begin{bmatrix} s[0] & \cdots & s[N-1] \end{bmatrix}^T \text{ and } \mathbf{x} = \begin{bmatrix} x[0] & \cdots & x[N-1] \end{bmatrix}^T$$

Linear Least Squares

Assume that an observed signal, $\{x[n]\}_0^{N-1}$, is a perturbed version of an unknown signal, $\{s[n]\}_0^{N-1}$, where each of these processes can be written by the random vectors:

$$\mathbf{s} = \begin{bmatrix} s[0] & \cdots & s[N-1] \end{bmatrix}^T \text{ and } \mathbf{x} = \begin{bmatrix} x[0] & \cdots & x[N-1] \end{bmatrix}^T$$

It is assumed the signal, $s[n]$, can be written as a linear combination of P known functions, $\{h_k[n]\}_{k=1}^P$, with weighting parameters $\{\theta_k\}_{k=1}^P$; thus:

$$s[n] = \sum_{k=1}^{P} \theta_k \, h_k[n]$$

Linear Least Squares

It is assumed the signal, $s[n]$, can be written as a linear combination of P known functions, $\{h_k[n]\}_{k=1}^P$, with weighting parameters $\{\theta_k\}_{k=1}^P$; thus:

$$s[n] = \sum_{k=1}^{P} \theta_k \, h_k[n]$$

Writing this in matrix-vector notation, it follows that:

$$\underbrace{\begin{bmatrix} s[0] \\ s[1] \\ \vdots \\ s[N-1] \end{bmatrix}}_{\mathbf{s}} = \underbrace{\begin{bmatrix} h_1[0] & h_2[0] & \cdots & h_P[0] \\ h_1[1] & h_2[1] & \cdots & h_P[1] \\ \vdots & \vdots & \ddots & \vdots \\ h_1[N-1] & h_2[N-1] & \cdots & h_P[N-1] \end{bmatrix}}_{\mathbf{H}} \underbrace{\begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_P \end{bmatrix}}_{\boldsymbol{\theta}}$$

Linear Least Squares

Writing this in matrix-vector notation, it follows that:

$$\underbrace{\begin{bmatrix} s[0] \\ s[1] \\ \vdots \\ s[N-1] \end{bmatrix}}_{\mathbf{s}} = \underbrace{\begin{bmatrix} h_1[0] & h_2[0] & \cdots & h_P[0] \\ h_1[1] & h_2[1] & \cdots & h_P[1] \\ \vdots & \vdots & \ddots & \vdots \\ h_1[N-1] & h_2[N-1] & \cdots & h_P[N-1] \end{bmatrix}}_{\mathbf{H}} \underbrace{\begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_P \end{bmatrix}}_{\boldsymbol{\theta}}$$

Thus, $\mathbf{s}$ is linear in the unknown parameter $\boldsymbol{\theta} = [\theta_1, \cdots, \theta_P]$:

$$\mathbf{s} = \mathbf{H}\,\boldsymbol{\theta}$$

Linear Least Squares

Writing this in matrix-vector notation, it follows that:

$$\underbrace{\begin{bmatrix} s[0] \\ s[1] \\ \vdots \\ s[N-1] \end{bmatrix}}_{\mathbf{s}} = \underbrace{\begin{bmatrix} h_1[0] & h_2[0] & \cdots & h_P[0] \\ h_1[1] & h_2[1] & \cdots & h_P[1] \\ \vdots & \vdots & \ddots & \vdots \\ h_1[N-1] & h_2[N-1] & \cdots & h_P[N-1] \end{bmatrix}}_{\mathbf{H}} \underbrace{\begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_P \end{bmatrix}}_{\boldsymbol{\theta}}$$

Thus, s is linear in the unknown parameter $\boldsymbol{\theta} = [\theta_1, \cdots, \theta_P]$:

$$\mathbf{s} = \mathbf{H}\,\boldsymbol{\theta}$$

The LSE is found by minimising:

$$J(\boldsymbol{\theta}) = \sum_{n=0}^{N-1} |x[n] - s[n]|^2 = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})$$

Linear Least Squares

Thus, s is linear in the unknown parameter $\boldsymbol{\theta} = [\theta_1, \cdots, \theta_P]$:

$$\mathbf{s} = \mathbf{H}\,\boldsymbol{\theta}$$

The LSE is found by minimising:

$$J(\boldsymbol{\theta}) = \sum_{n=0}^{N-1} |x[n] - s[n]|^2 = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})$$

This can be written as:

$$J(\boldsymbol{\theta}) = \mathbf{x}^T\mathbf{x} - 2\mathbf{x}^T\mathbf{H}\boldsymbol{\theta} + \boldsymbol{\theta}^T\mathbf{H}^T\mathbf{H}\boldsymbol{\theta}$$

Linear Least Squares

The LSE is found by minimising:

$$J(\boldsymbol{\theta}) = \sum_{n=0}^{N-1} |x[n] - s[n]|^2 = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})$$

This can be written as:

$$J(\boldsymbol{\theta}) = \mathbf{x}^T\mathbf{x} - 2\mathbf{x}^T\mathbf{H}\boldsymbol{\theta} + \boldsymbol{\theta}^T\mathbf{H}^T\mathbf{H}\boldsymbol{\theta}$$

and using the two identities that:

$$\frac{\partial \mathbf{b}^T\mathbf{a}}{\partial \mathbf{a}} = \mathbf{b} \quad \text{and} \quad \frac{\partial \mathbf{a}^T\mathbf{B}\mathbf{a}}{\partial \mathbf{a}} = \left(\mathbf{B} + \mathbf{B}^T\right)\mathbf{a}$$

Linear Least Squares

The LSE is found by minimising:

$$J(\boldsymbol{\theta}) = \sum_{n=0}^{N-1} |x[n] - s[n]|^2 = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})$$

This can be written as:

$$J(\boldsymbol{\theta}) = \mathbf{x}^T\mathbf{x} - 2\mathbf{x}^T\mathbf{H}\boldsymbol{\theta} + \boldsymbol{\theta}^T\mathbf{H}^T\mathbf{H}\boldsymbol{\theta}$$

and using the two identities that:

$$\frac{\partial \mathbf{b}^T\mathbf{a}}{\partial \mathbf{a}} = \mathbf{b} \quad \text{and} \quad \frac{\partial \mathbf{a}^T\mathbf{B}\mathbf{a}}{\partial \mathbf{a}} = \left(\mathbf{B} + \mathbf{B}^T\right)\mathbf{a}$$

then observing in this case $\mathbf{B} = \mathbf{H}^T\mathbf{H} = \mathbf{B}^T$ it follows that

$$\frac{\partial J(\boldsymbol{\theta})}{\partial \theta} = -2\mathbf{H}^T\mathbf{x} + 2\mathbf{H}^T\mathbf{H}\theta$$

Linear Least Squares

Setting the gradient of $J(\boldsymbol{\theta})$ to zero yields the LSE:

$$\hat{\boldsymbol{\theta}}_{LSE} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x}$$

Linear Least Squares

Setting the gradient of $J(\boldsymbol{\theta})$ to zero yields the LSE:

$$\hat{\boldsymbol{\theta}}_{LSE} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x}$$

- The equations $\mathbf{H}^T\mathbf{H}\boldsymbol{\theta} = \mathbf{H}^T\mathbf{x}$, to be solved for $\hat{\boldsymbol{\theta}}$, are termed the **normal equation**.

Linear Least Squares

Setting the gradient of $J(\boldsymbol{\theta})$ to zero yields the LSE:

$$\hat{\boldsymbol{\theta}}_{LSE} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x}$$

- The equations $\mathbf{H}^T\mathbf{H}\boldsymbol{\theta} = \mathbf{H}^T\mathbf{x}$ are the **normal equation**.
- Requiring $\mathbf{H}$ to be full rank guarantees invertibility of $\mathbf{H}^T\mathbf{H}$.

Linear Least Squares

Setting the gradient of $J(\boldsymbol{\theta})$ to zero yields the LSE:

$$\hat{\boldsymbol{\theta}}_{LSE} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x}$$

- The equations $\mathbf{H}^T\mathbf{H}\boldsymbol{\theta} = \mathbf{H}^T\mathbf{x}$ are the **normal equation.**

- The minimum LS error is found from:

$$J_{\min} = J(\hat{\boldsymbol{\theta}}) = \left(\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}}\right)^T\left(\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}}\right)$$
$$= \left(\mathbf{x} - \mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x}\right)^T\left(\mathbf{x} - \mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x}\right)$$

$$J_{\min} = \mathbf{x}^T\left(\mathbf{I}_N - \mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\right)\left(\mathbf{I}_N - \mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\right)\mathbf{x}$$

Linear Least Squares

Setting the gradient of $J(\boldsymbol{\theta})$ to zero yields the LSE:

$$\hat{\boldsymbol{\theta}}_{LSE} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x}$$

- The equations $\mathbf{H}^T\mathbf{H}\boldsymbol{\theta} = \mathbf{H}^T\mathbf{x}$ are the **normal equation.**
- The minimum LS error is found from:

$$J_{\min} = \mathbf{x}^T\left(\mathbf{I}_N - \mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\right)\left(\mathbf{I}_N - \mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\right)\mathbf{x}$$

$\mathbf{A} = \mathbf{I}_N - \mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T$ is **idempotent** so $\mathbf{A}^2 = \mathbf{A}$. Hence:

$$J_{\min} = \mathbf{x}^T\left(\mathbf{I}_N - \mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\right)\mathbf{x}$$

Linear Least Squares

Example (Fourier Series Estimation). An application of the general linear model is in spectral estimation. Suppose that a signal, $s[n]$, is modelled as the sum of sinusoids:

$$s[n] = \sum_{p=1}^{P} a_p \sin(p\,\omega_0\,n) + b_p \cos(p\omega_0 n) \qquad \bowtie$$

where $\{a_p,\, b_p\}_{p=1}^{P}$ are the unknown amplitudes to be estimated, and the fundamental, ω_0, and model order P, are assumed to be known.

The signal, $s[n]$, is observed in noise. Write down the least squares solution.

Linear Least Squares

Example (Fourier Series Estimation). Suppose that :

$$s[n] = \sum_{p=1}^{P} a_p \sin(p\,\omega_0\,n) + b_p \cos(p\omega_0 n)$$

where $\{a_p,\, b_p\}_{p=1}^{P}$ are the unknown amplitudes to be estimated.

SOLUTION. Writing the relationship between the observation, signal model, and modelling error:

$$x[n] = s[n] + e[n] = \sum_{p=1}^{P} (a_p \sin\omega_p\,n + b_p \cos\omega_p n) + e[n]$$

□

Linear Least Squares

Example (Fourier Series Estimation). Suppose that :

$$s[n] = \sum_{p=1}^{P} a_p \sin(p\,\omega_0\, n) + b_p \cos(p\omega_0 n)$$

where $\{a_p,\, b_p\}_{p=1}^{P}$ are the unknown amplitudes to be estimated.

SOLUTION. This model can be written in a linear in the parameters (LITP) form by defining, where $\ell \triangleq N-1$:

$$\mathbf{H} = \begin{bmatrix} 0 & 1 & 0 & 1 & \cdots & 0 & 1 \\ \sin\omega_0 & \cos\omega_0 & \sin 2\omega_0 & \cos 2\omega_0 & \cdots & \sin P\omega_0 & \cos P\omega_0 \\ \sin 2\omega_0 & \cos 2\omega_0 & \sin 4\omega_0 & \cos 4\omega_0 & \cdots & \sin 2P\omega_0 & \cos 2P\omega_0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \sin \ell\omega_0 & \cos \ell\omega_0 & \sin 2\ell\omega_0 & \cos 2\ell\omega_0 & \cdots & \sin P\ell\omega_0 & \cos P\ell\omega_0 \end{bmatrix}$$

Linear Least Squares

Example (Fourier Series Estimation). Suppose that :

$$s[n] = \sum_{p=1}^{P} a_p \sin\left(p\,\omega_0\,n\right) + b_p \cos\left(p\omega_0 n\right)$$

where $\{a_p,\, b_p\}_{p=1}^{P}$ are the unknown amplitudes to be estimated.

SOLUTION. Hence, with the parameter vector defined as:

$$\boldsymbol{\theta} = \begin{bmatrix} a_1 & b_1 & a_2 & b_2 & \cdots & a_P & b_P \end{bmatrix}^T$$

the signal model is $\mathbf{s} = \mathbf{H}\boldsymbol{\theta}$, and the linear LSE estimator is:

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x} \qquad \square$$

where $\hat{\boldsymbol{\theta}}$ is of dimension $2P$, and therefore $\mathbf{H}$ is $N \times 2P$.

Linear Least Squares

Example (Fourier Series Estimation). Suppose that :

$$s[n] = \sum_{p=1}^{P} a_p \sin(p\,\omega_0\,n) + b_p \cos(p\omega_0 n)$$

where $\{a_p,\, b_p\}_{p=1}^{P}$ are the unknown amplitudes to be estimated.

SOLUTION. Hence, with the parameter vector defined as:

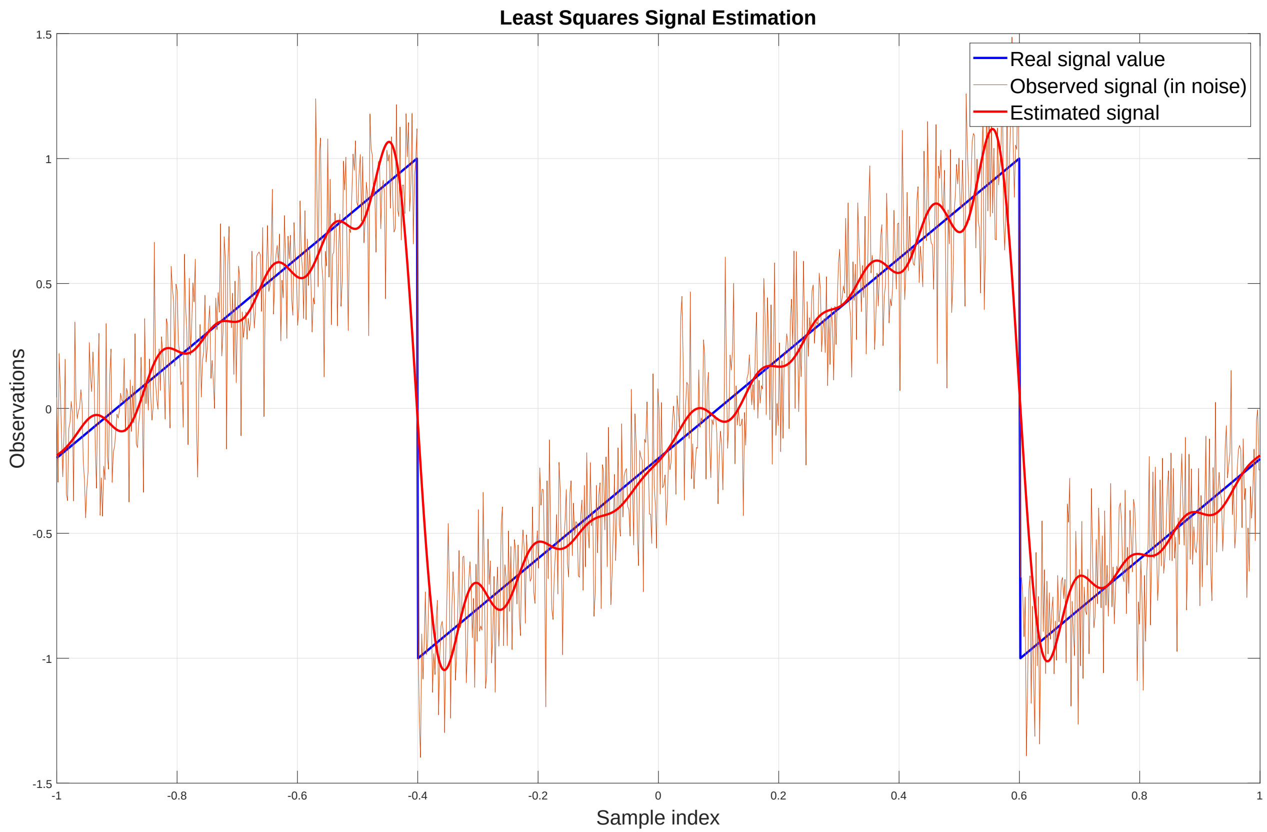
$$\boldsymbol{\theta} = \begin{bmatrix} a_1 & b_1 & a_2 & b_2 & \cdots & a_P & b_P \end{bmatrix}^T$$

the signal model is $\mathbf{s} = \mathbf{H}\boldsymbol{\theta}$, and the linear LSE estimator is:

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x} \qquad \square$$

Using the orthognality of the Fourier basis, this can simplify .

Linear Least Squares



In this figure, the true underlying signal model is shown (the sawtooth), the observed signal (with sensor noise), and the estimated Fourier signal model.

Linear Least Squares

– End-of-Topic 43: **Introduction to Linear Least Squares Estimation –**

Any Questions?

Bayesian Parameter Estimation

Using the method of maximum likelihood (or least squares) to infer the values of a parameter has significant limitations:

1. - First, the likelihood function does not use information other than the data itself to *infer* the values of the parameters.
   - No prior knowledge, stated before the data is observed, is utilised regarding the possible or probable values that the parameters might take.
   - In many applications, a physical understanding of the problem at hand, or of the circumstances surrounding how an experiment is conducted, can suggest that some values of the parameters are impossible, and that some are more likely to occur than others.

Bayesian Parameter Estimation

Using the method of maximum likelihood (or least squares) to infer the values of a parameter has significant limitations:

1. The likelihood function on its own does not limit the number of parameters in a model used to fit the data. The number of parameters is chosen in advance, by the Signal Processing Engineer, but the likelihood function does not indicate whether the number of parameters chosen is more than necessary to model the data, or less than needed.

Bayesian Parameter Estimation

– End-of-Topic 44: **Introduction to Advanced Bayesian Parameter Estimation** –

Any Questions?

Lecture Slideset 5
MonteCarlo

Introduction

Many signal processing problems can be reduced to either an *optimisation* problem or an *integration* problem:

Introduction

Many signal processing problems can be reduced to either an *optimisation* problem or an *integration* problem:

Optimisation: involves finding the solution to

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\Theta} h(\boldsymbol{\theta})$$

where $h(\cdot)$ is a scalar function of a multi-dimensional vector of parameters, $\boldsymbol{\theta}$.

- Typically, $h(\cdot)$ might represent some **cost function**, and it is implicitly assumed that the optimisation cannot be calculated explicitly.

Introduction

Many signal processing problems can be reduced to either an *optimisation* problem or an *integration* problem:

Integration: involves evaluating an integral,

$$\mathcal{I} = \int_{\Theta} f(\boldsymbol{\theta})\, d\boldsymbol{\theta},$$

that cannot explicitly be calculated in *closed form*.

Introduction

Many signal processing problems can be reduced to either an *optimisation* problem or an *integration* problem:

Integration: involves evaluating an integral,

$$\mathcal{I} = \int_{\Theta} f(\boldsymbol{\theta})\, d\boldsymbol{\theta},$$

that cannot explicitly be calculated in *closed form*.

- For example, the Gaussian-error function:

$$\Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{\theta^2}{2}}\, d\theta$$

Again, the integral may be multi-dimensional, and in general $\boldsymbol{\theta}$ is a vector.

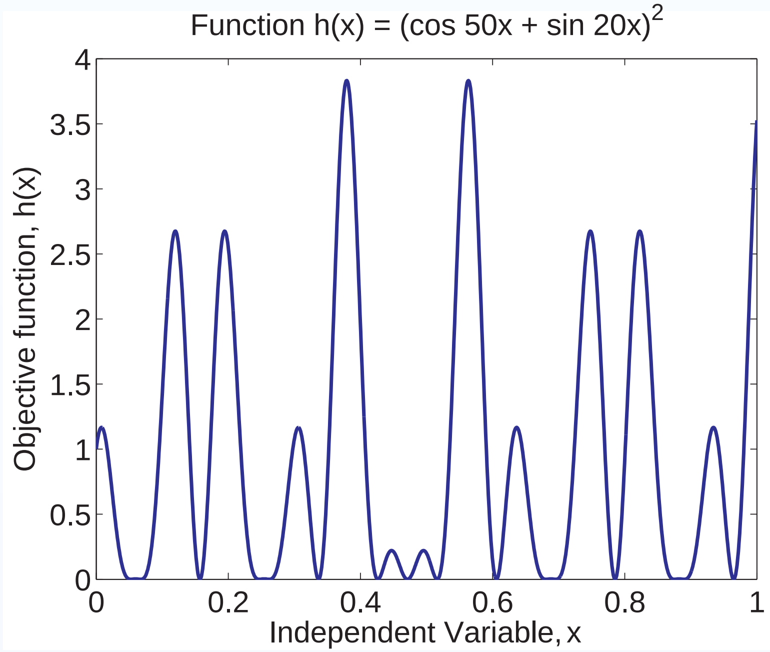
Introduction

Many signal processing problems can be reduced to either an *optimisation* problem or an *integration* problem:

Optimisation and Integration Some problems involve both integration and optimisation: a fundamental problem is the maximisation of a marginal distribution:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta} \in \Theta} \int_{\Omega} f(\boldsymbol{\theta}, \boldsymbol{\omega}) \, d\boldsymbol{\omega}$$

Deterministic Numerical Methods



Plot of the function $h(x) = (\cos 50x + \sin 20x)^2$, $0 \leq x \leq 1$.

There are various deterministic solutions to the optimisation and integration problems.

Deterministic Numerical Methods

Optimisation:

1. Golden-section search and Brent's Method in one dimension;
2. Nelder and Mead Downhill Simplex method in multi-dimensions;
3. Gradient and Variable-Metric methods in multi-dimensions, typically an extension of Newton-Raphson methods.

Deterministic Numerical Methods

Integration: Most deterministic integration rely on classic formulas for equally spaced abscissas:

1. simple Riemann integration;
2. standard and extended Simpson's and Trapezoidal rules;
3. refinements such as Romberg Integration.

Deterministic Numerical Methods

Integration: Most deterministic integration rely on classic formulas for equally spaced abscissas:

1. simple Riemann integration;
2. standard and extended Simpson's and Trapezoidal rules;
3. refinements such as Romberg Integration.

More sophisticated approaches allow non-uniformally spaced abscissas at which the function is evaluated.

- These methods tend to use Gaussian quadratures and orthogonal polynomials. Splines are also used.

Deterministic Numerical Methods

Integration: Most deterministic integration rely on classic formulas for equally spaced abscissas:

1. simple Riemann integration;
2. standard and extended Simpson's and Trapezoidal rules;
3. refinements such as Romberg Integration.

More sophisticated approaches allow non-uniformally spaced abscissas at which the function is evaluated.

- These methods tend to use Gaussian quadratures and orthogonal polynomials. Splines are also used.

Unfortunately, these methods are not easily extended to multi-dimensions.

Deterministic Optimisation

The **Nelder-Mead Downhill Simplex method** simply crawls downhill in a straightforward fashion that makes almost no special assumptions about your function.

- This can be extremely slow, but it can be robust.

Deterministic Optimisation

The **Nelder-Mead Downhill Simplex method** simply crawls downhill in a straightforward fashion that makes almost no special assumptions about your function.

- This can be extremely slow, but it can be robust.

Gradient methods are typically based on the Newton-Raphson algorithm which solves $\nabla h(\boldsymbol{\theta}) = \mathbf{0}$.

- For a scalar function, $h(\boldsymbol{\theta})$, of a vector of independent variables $\boldsymbol{\theta}$, a sequence $\boldsymbol{\theta}_n$ is produced such that:

Deterministic Optimisation

The **Nelder-Mead Downhill Simplex method** simply crawls downhill in a straightforward fashion that makes almost no special assumptions about your function.

- This can be extremely slow, but it can be robust.

Gradient methods are typically based on the Newton-Raphson algorithm which solves $\nabla h(\boldsymbol{\theta}) = \mathbf{0}$.

- For a scalar function, $h(\boldsymbol{\theta})$, of a vector of independent variables $\boldsymbol{\theta}$, a sequence $\boldsymbol{\theta}_n$ is produced such that:

$$\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n - \left(\nabla\,\nabla^T h\left(\boldsymbol{\theta}_n\right)\right)^{-1} \nabla h\left(\boldsymbol{\theta}_n\right)$$

Numerous variants of Newton-Raphson-type techniques exist, and include the **steepest descent method**, or the **Levenberg-Marquardt method**.

Deterministic Integration

Assuming θ is a scalar and $b > a$, the integral

$$\mathcal{I} = \int_a^b f(\theta)\, d\theta,$$

can be solved with the trapezoidal rule:

$$\hat{I} = \frac{1}{2} \sum_{k=0}^{N-1} (\theta_{k+1} - \theta_k)\left(f(\theta_k) + f(\theta_{k+1})\right)$$

where the θ_k's constitute an ordered partition of $[a,\ b]$.

Deterministic Integration

Assuming θ is a scalar and $b > a$, the integral

$$\mathcal{I} = \int_a^b f(\theta)\, d\theta,$$

can be solved with the trapezoidal rule:

$$\hat{I} = \frac{1}{2} \sum_{k=0}^{N-1} (\theta_{k+1} - \theta_k)\,(f(\theta_k) + f(\theta_{k+1}))$$

where the θ_k's constitute an ordered partition of $[a,\, b]$.

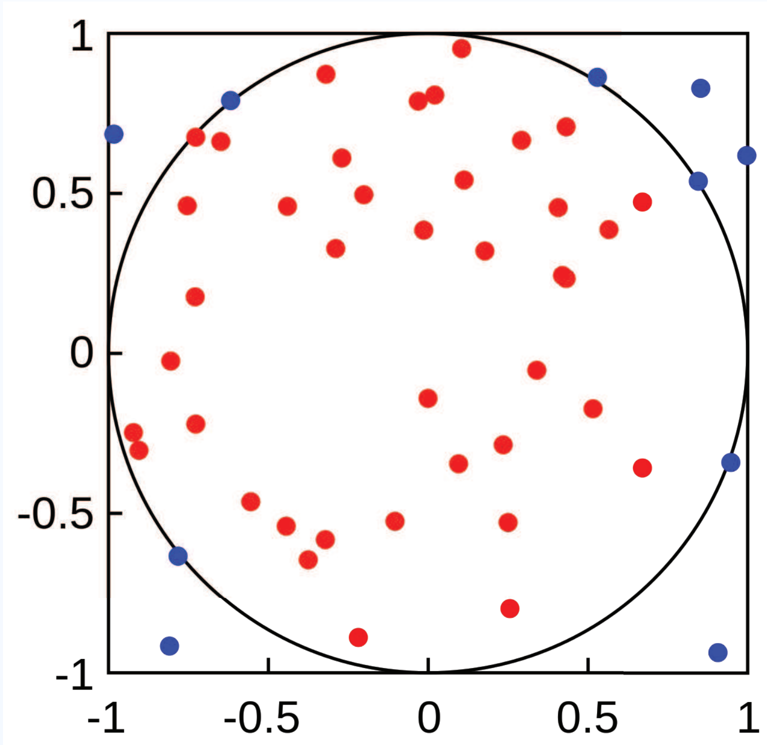
- Another formula is **Simpson's rule**:

$$\hat{I} = \frac{\delta}{3} \left\{ f(a) + 4 \sum_{k=1}^{N} f(\theta_{2k-1}) + 2 \sum_{k=1}^{N} h(\theta_{2k}) + f(b) \right\}$$

in the case of equally spaced samples with $\delta = \theta_{k+1} - \theta_k$.

Monte Carlo Numerical Methods

Monte Carlo methods are stochastic techniques, in which random numbers are generated and use to examine some problem.



Estimating the value of π through Monte Carlo integration.

Monte Carlo Integration

Consider the integral,

$$\mathcal{I} = \int_{\Theta} f(\boldsymbol{\theta})\, d\boldsymbol{\theta}.$$

Monte Carlo Integration

$$\mathcal{I} = \int_\Theta f(\boldsymbol{\theta})\, d\boldsymbol{\theta}.$$

Defining a function $\pi(\boldsymbol{\theta})$ which is non-zero and positive for all $\boldsymbol{\theta} \in \Theta$, this integral can be expressed in the alternate form:

$$\mathcal{I} = \int_\Theta \frac{f(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta})}\, \pi(\boldsymbol{\theta})\, d\boldsymbol{\theta},$$

where the function $\pi(\boldsymbol{\theta}) > 0$, $\boldsymbol{\theta} \in \Theta$ is a pdf which satisfies

$$\int_\Theta \pi(\boldsymbol{\theta})\, d\boldsymbol{\theta} = 1$$

Monte Carlo Integration

$$\mathcal{I} = \int_{\Theta} f(\boldsymbol{\theta})\, d\boldsymbol{\theta}.$$

Defining a function $\pi(\boldsymbol{\theta})$ which is non-zero and positive for all $\boldsymbol{\theta} \in \Theta$, this integral can be expressed in the alternate form:

$$\mathcal{I} = \int_{\Theta} \frac{f(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta})}\, \pi(\boldsymbol{\theta})\, d\boldsymbol{\theta},$$

where the function $\pi(\boldsymbol{\theta}) > 0$, $\boldsymbol{\theta} \in \Theta$ is a pdf which satisfies

$$\int_{\Theta} \pi(\boldsymbol{\theta})\, d\boldsymbol{\theta} = 1$$

This may be written as an expectation:

$$\mathcal{I} = \mathbb{E}_{\pi}\left[\frac{f(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta})}\right]$$

Monte Carlo Integration

This expectation can be estimated using the idea of the **sample expectation**, and leads to the idea behind Monte Carlo integration:

1. Sample N random variates from a density function $\pi(\boldsymbol{\theta})$,

$$\boldsymbol{\theta}^{(k)} \sim \pi(\boldsymbol{\theta}), \quad k \in \mathcal{N} = \{0, \dots, N-1\}$$

2. Calculate the sample average of the expectation using

$$\hat{\mathcal{I}} = \frac{1}{N} \sum_{k=0}^{N-1} \frac{f(\boldsymbol{\theta}^{(k)})}{\pi(\boldsymbol{\theta}^{(k)})} \approx \mathbb{E}_\pi \left[\frac{f(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta})} \right]$$

Stochastic Optimisation

There are two distinct approaches to the Monte Carlo optimisation of the objective function $h(\boldsymbol{\theta})$:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta} \in \Theta} h(\boldsymbol{\theta})$$

The first method is broadly known as an **exploratory approach**, while the second approach is based on a **probabilistic approximation** of the objective function.

Stochastic Optimisation

Exploratory approach This approach is concerned with fast *explorations* of the sample space rather than working with the objective function directly.

Stochastic Optimisation

Exploratory approach This approach is concerned with fast *explorations* of the sample space rather than working with the objective function directly.

For example, maximisation can be solved by sampling a large number, N, of independent random variables, $\{\boldsymbol{\theta}^{(k)}\}$, from a pdf $\pi(\boldsymbol{\theta})$, and taking the estimate:

$$\hat{\boldsymbol{\theta}} \approx \arg\max_{\{\boldsymbol{\theta}^{(k)}\}} h\left(\boldsymbol{\theta}^{(k)}\right)$$

Typically, when no specific features regarding the function $h\left(\boldsymbol{\theta}\right)$, are taken into account, $\pi(\boldsymbol{\theta})$ will take on a uniform distribution over Θ.

Stochastic Optimisation

Exploratory approach This approach is concerned with fast *explorations* of the sample space rather than working with the objective function directly.

For example, maximisation can be solved by sampling a large number, N, of independent random variables, $\{\boldsymbol{\theta}^{(k)}\}$, from a pdf $\pi(\boldsymbol{\theta})$, and taking the estimate:

$$\hat{\boldsymbol{\theta}} \approx \arg\max_{\{\boldsymbol{\theta}^{(k)}\}} h\left(\boldsymbol{\theta}^{(k)}\right)$$

Typically, when no specific features regarding the function $h(\boldsymbol{\theta})$, are taken into account, $\pi(\boldsymbol{\theta})$ will take on a uniform distribution over Θ.

Stochastic Approximation The Monte Carlo EM algorithm

Generating Random Variables

This section discusses a variety of techniques for generating random variables from a different distributions.

Uniform Variates

The foundation underpinning all stochastic simulations is the ability to generate a sequence of i. i. d. uniform random variates over the range $(0, 1]$.

Uniform Variates

The foundation underpinning all stochastic simulations is the ability to generate a sequence of i. i. d. uniform random variates over the range $(0, 1]$.

Random variates are *pseudo* or *synthetic* and not truly random since they are usually generated using a recurrence of the form:

$$x_{n+1} = (a\, x_n + b) \mod m$$

Uniform Variates

The foundation underpinning all stochastic simulations is the ability to generate a sequence of i. i. d. uniform random variates over the range $(0, 1]$.

Random variates are *pseudo* or *synthetic* and not truly random since they are usually generated using a recurrence of the form:

$$x_{n+1} = (a\, x_n + b) \mod m$$

This is known as the linear congruential generator.

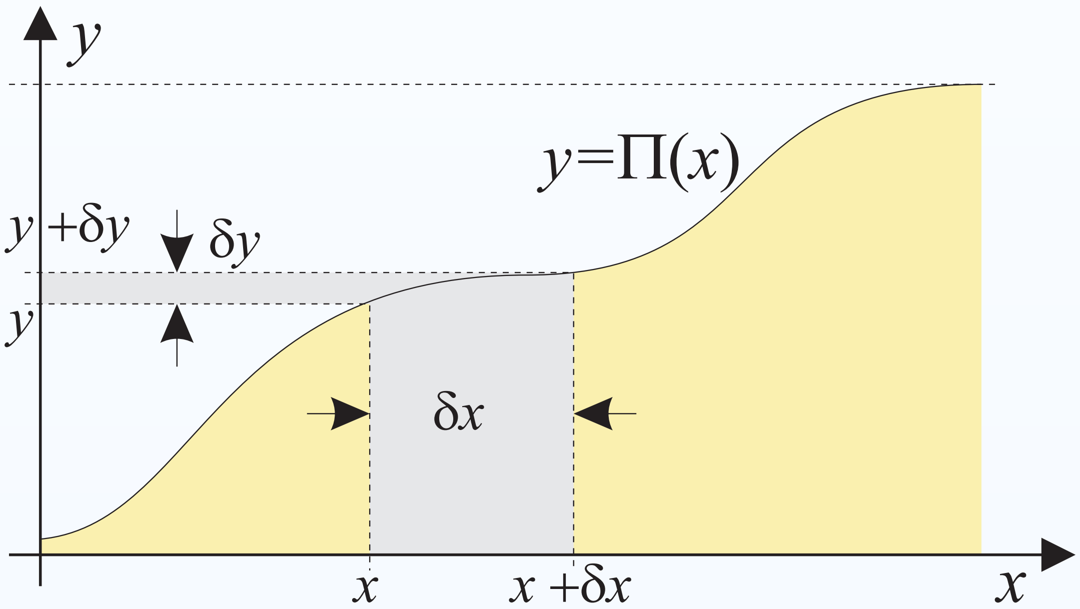
However, suitable values of a, b and m can be chosen such that the random variates pass all statistical tests of randomness.

Transformation Methods

It is possible to sample from a number of extremely important probability distributions by applying various probability transformation methods.

Theorem (Probability transformation rule). PROOF. The proof is given in the handout on scalar random variables.

Inverse Transform Method

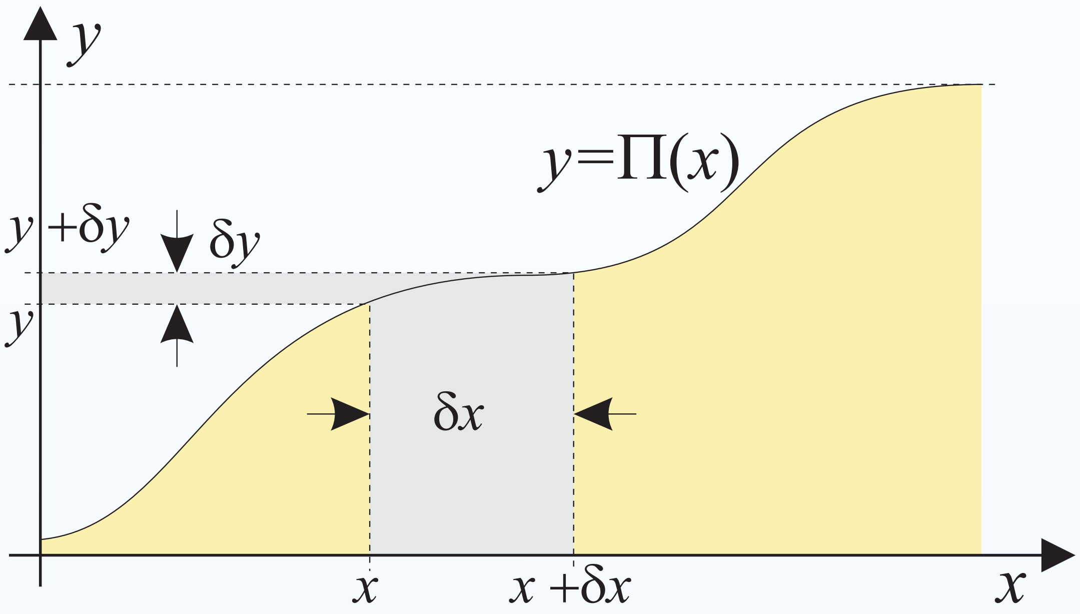


A simple derivation of the inverse transform method

$X(\zeta)$ and $Y(\zeta)$ are RVs related by the function $Y(\zeta) = \Pi(X(\zeta))$.

- $\Pi(\zeta)$ is monotonically increasing so that there is only one solution to the equation $y = \Pi(x)$, $x = \Pi^{-1}(y)$.

Inverse Transform Method



A simple derivation of the inverse transform method

$$f_X(x) = \frac{d\Pi(x)}{dx} f_Y(y)$$

Now, suppose $Y(\zeta) \sim \mathcal{U}_{[0,1]}$ is a uniform random variable. If $\Pi(x)$ is the cdf corresponding to a desired pdf $\pi(x)$, then

$$f_X(x) = \pi(x), \quad \text{where} \quad x = \Pi^{-1}(y)$$

Inverse Transform Method

In otherwords, if

$$U(\zeta) \sim \mathcal{U}_{[0,\,1]},\ X(\zeta) = \Pi^{-1} U(\zeta) \sim \pi\left(x\right)$$

Inverse Transform Method

In otherwords, if

$$U(\zeta) \sim \mathcal{U}_{[0,\,1]},\; X(\zeta) = \Pi^{-1} U(\zeta) \sim \pi\left(x\right)$$

Example (Exponential variable generation). If $X(\zeta) \sim \mathcal{E}xp(1)$, such that $\pi(x) = e^{-x}$ and $\Pi(x) = 1 - e^{-x}$, then solving for x in terms of $u = 1 - e^{-x}$ gives $x = -\log(1-u)$.

Inverse Transform Method

In otherwords, if

$$U(\zeta) \sim \mathcal{U}_{[0,\,1]},\; X(\zeta) = \Pi^{-1} U(\zeta) \sim \pi\,(x)$$

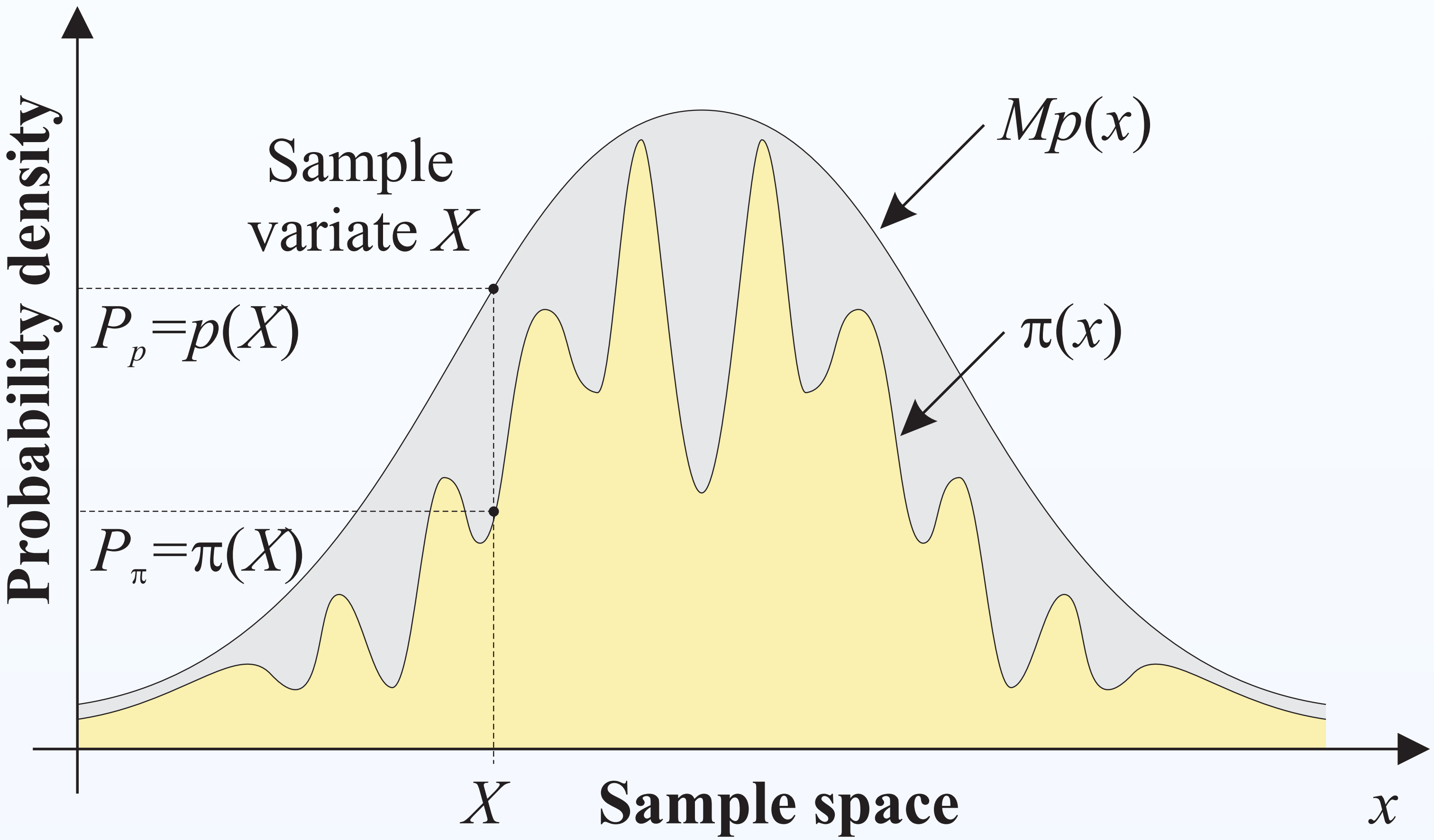
Example (Exponential variable generation). If $X(\zeta) \sim \mathcal{E}xp(1)$, such that $\pi(x) = e^{-x}$ and $\Pi(x) = 1 - e^{-x}$, then solving for x in terms of $u = 1 - e^{-x}$ gives $x = -\log(1-u)$.

- Therefore, if $U(\zeta) \sim \mathcal{U}_{[0,\,1]}$, then the RV from the transformation $X(\zeta) = -\log U(\zeta)$ has the exponential distribution (since $U(\zeta)$ and $1 - U(\zeta)$ are both uniform). $\bowtie$

Acceptance-Rejection Sampling

For most distributions, it is often difficult or even impossible to directly simulate using either the inverse transform or probability transformations.

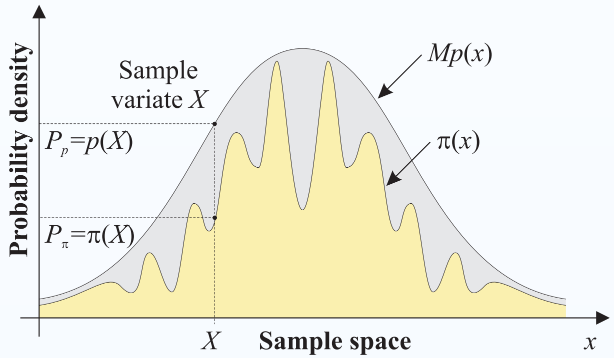
Acceptance-Rejection Sampling



On average, you would expect to have too many variates that take on the value X by a factor of

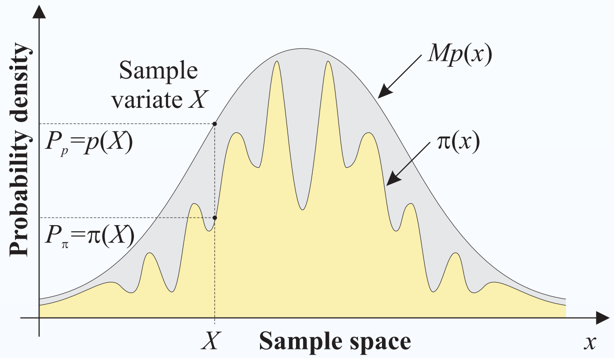
$$u(X) = \frac{P_p}{P_\pi} = \frac{p(X)}{\pi(X)}$$

Acceptance-Rejection Sampling



Thus, to reduce the number of variates that take on a value of X, simply throw away a number of samples in proportion to the amount of *over sampling*.

Acceptance-Rejection Sampling



Thus, to reduce the number of variates that take on a value of X, simply throw away a number of samples in proportion to the amount of *over sampling*.

1. Generate the random variates $X \sim p(x)$ and $U \sim \mathcal{U}_{[0,\,1]}$;
2. Accept X if $U \leq P_a = \frac{\pi(X)}{M p(x)}$;
3. Otherwise, reject and return to first step.

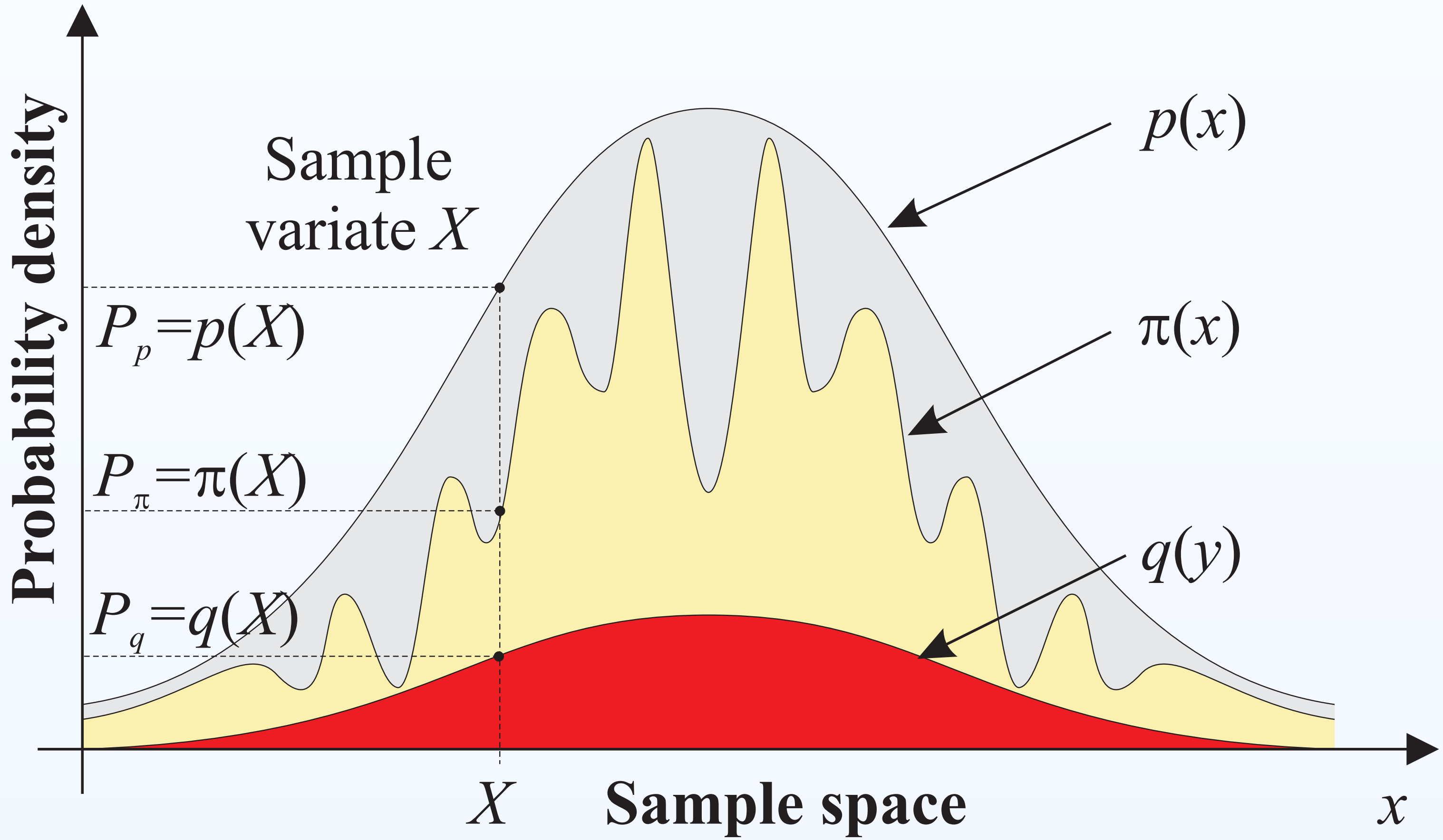
Envelope and Squeeze Methods

A problem with many sampling methods, which can make the density $\pi(x)$ difficult to simulate, is that the function may require substantial computing time at each evaluation.

It is possible to reduce the algorithmic complexity by looking for another computationally simple function, $q(x)$ which *bounds* $\pi(x)$ *from below*.

Envelope and Squeeze Methods

If X satisfies $q(X) \leq \pi(X)$, then it should be accepted when $U \leq \frac{q(X)}{Mp(x)}$, since this also satisfies $U \leq \frac{\pi(X)}{Mp(x)}$.



Envelope and Squeeze Methods

This leads to the **envelope accept-reject algorithm**:

1. Generate the random variates $X \sim p(x)$ and $U \sim \mathcal{U}_{[0,\,1]}$;
2. Accept X if $U \le \frac{q(X)}{M p(x)}$;
3. Otherwise, accept X if $U \le \frac{\pi(X)}{M p(x)}$;
4. Otherwise, reject and return to first step.

Envelope and Squeeze Methods

This leads to the **envelope accept-reject algorithm**:

1. Generate the random variates $X \sim p(x)$ and $U \sim \mathcal{U}_{[0,\,1]}$;
2. Accept X if $U \leq \frac{q(X)}{M p(x)}$;
3. Otherwise, accept X if $U \leq \frac{\pi(X)}{M p(x)}$;
4. Otherwise, reject and return to first step.

By construction of a lower envelope on $\pi(x)$, the number of function evaluations is potentially decreased by a factor of

$$P_{\bar{\pi}} = \frac{1}{M} \int q(x)\, dx$$

which is the probability that $\pi(x)$ is not evaluated.

Importance Sampling

The problem with accept-reject sampling methods is finding the envelope functions and the constant M.

Importance Sampling

The problem with accept-reject sampling methods is finding the envelope functions and the constant M.

The simplest application of **importance sampling** is in Monte Carlo integration. Suppose that is is desired to evaluate the function:

$$\mathcal{I} = \int_{\Theta} f(\boldsymbol{\theta})\, d\boldsymbol{\theta}.$$

Importance Sampling

The problem with accept-reject sampling methods is finding the envelope functions and the constant M.

The simplest application of **importance sampling** is in Monte Carlo integration. Suppose that is is desired to evaluate the function:

$$\mathcal{I} = \int_{\Theta} f(\boldsymbol{\theta})\, d\boldsymbol{\theta}.$$

Approximate by empirical average:

$$\hat{\mathcal{I}} = \frac{1}{N} \sum_{k=0}^{N-1} \mathbb{I}_{\Theta}\left(\boldsymbol{\theta}^{(k)}\right), \quad \text{where} \quad \boldsymbol{\theta}^{(k)} \sim f(\boldsymbol{\theta})$$

where $\mathbb{I}_{\mathcal{A}}(a)$ is the indicator function, and is equal to one if $a \in \mathcal{A}$ and zero otherwise.

Importance Sampling

Defining an *easy-to-sample-from* density $\pi(\boldsymbol{\theta}) > 0, \forall \boldsymbol{\theta} \in \Theta$:

$$\mathcal{I} = \int_{\Theta} \frac{f(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta})} \pi(\boldsymbol{\theta}) \, d\boldsymbol{\theta} = \mathbb{E}_{\pi} \left[\frac{f(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta})} \right],$$

Importance Sampling

Defining an *easy-to-sample-from* density $\pi(\boldsymbol{\theta}) > 0, \forall \boldsymbol{\theta} \in \Theta$:

$$\mathcal{I} = \int_{\Theta} \frac{f(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta})} \pi(\boldsymbol{\theta}) \, d\boldsymbol{\theta} = \mathbb{E}_{\pi}\left[\frac{f(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta})}\right],$$

leads to an estimator based on the **sample expectation**;

$$\hat{\mathcal{I}} = \frac{1}{N} \sum_{k=0}^{N-1} \frac{f(\boldsymbol{\theta}^{(k)})}{\pi(\boldsymbol{\theta}^{(k)})}$$

Other Methods

Include:

- representing pdfs as mixture of distributions;
- algorithms for log-concave densities, such as the adaptive rejection sampling scheme;
- generalisations of accept-reject;
- method of composition (similar to Gibbs sampling);
- ad-hoc methods, typically based on probability transformations and order statistics (for example, generating Beta distributions with integer parameters).

Markov chain Monte Carlo Methods

A **Markov chain** is the first generalisation of an independent process, where each *state* of a Markov chain depends on the previous state only.

The Metropolis-Hastings algorithm

The **Metropolis-Hastings algorithm** is an extremely flexible method for producing a random sequence of samples from a given density.

1. Generate a random sample from a **proposal distribution**: $Y \sim g\left(y \,|\, X^{(k)}\right)$.

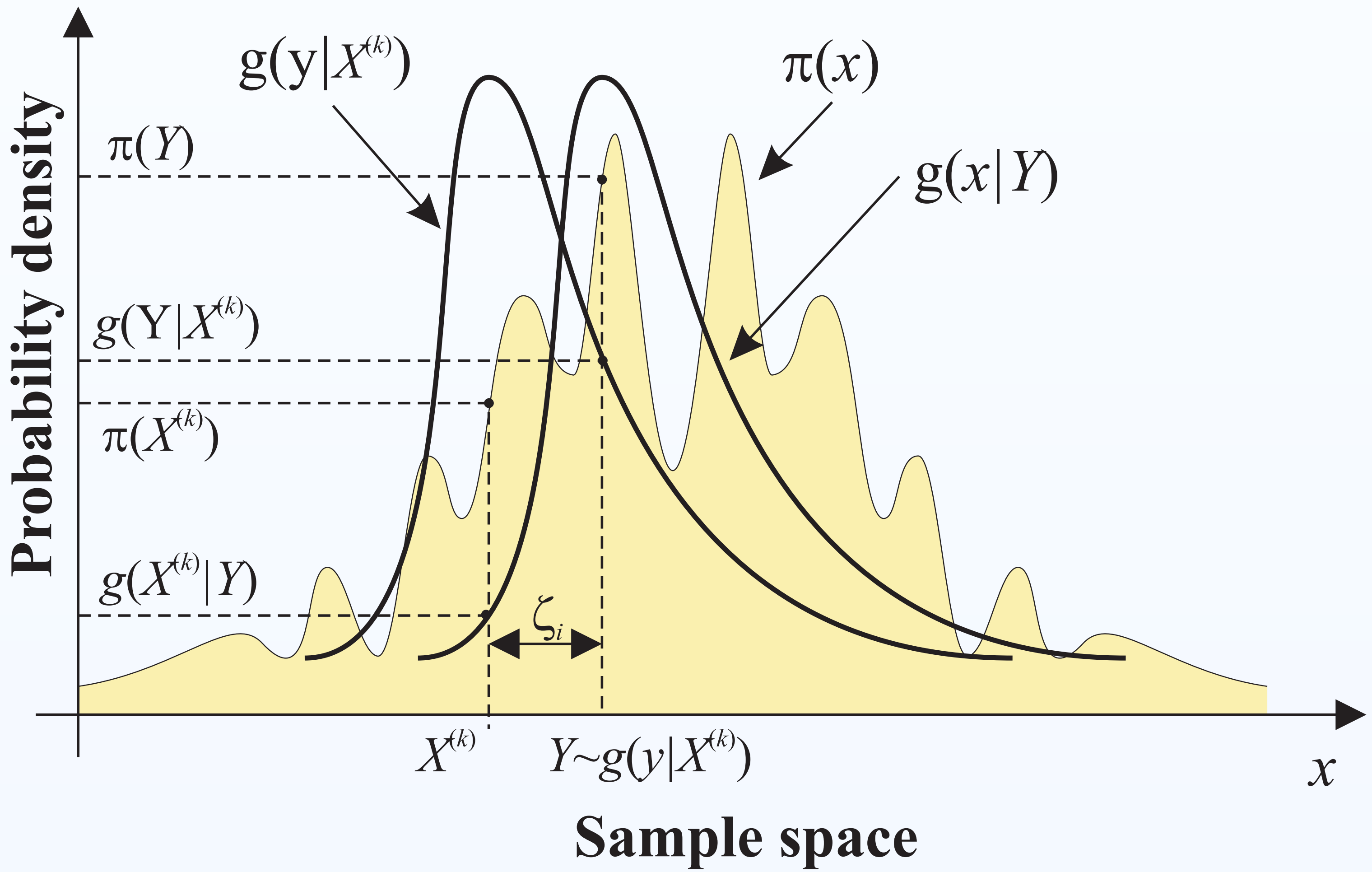
2. Set the new random variate to be:

$$X^{(k+1)} = \begin{cases} Y & \text{with probability } \rho(X^{(k)}, Y) \\ X^{(k)} & \text{with probability } 1 - \rho(X^{(k)}, Y) \end{cases}$$

where the acceptance ratio function $\rho(x, y)$ is given by:

$$\rho(x, y) = \min\left\{ \frac{\pi(y)}{g(y \,|\, x)} \left(\frac{\pi(x)}{g(x \,|\, y)} \right)^{-1}, 1 \right\} \equiv \min\left\{ \frac{\pi(y)}{\pi(x)} \frac{g(x \,|\, y)}{g(y \,|\, x)}, 1 \right\}$$

The Metropolis-Hastings algorithm



Gibbs Sampling

Gibbs sampling is a Monte Carlo method that facilitates sampling from a multivariate density function, $\pi\left(\theta_0, \theta_1, \ldots, \theta_M\right)$ by drawing successive samples from marginal densities of smaller dimensions.

Gibbs Sampling

Gibbs sampling is a Monte Carlo method that facilitates sampling from a multivariate density function, $\pi\left(\theta_0, \theta_1, \dots, \theta_M\right)$ by drawing successive samples from marginal densities of smaller dimensions.

Using the probability chain rule,

$$\pi\left(\{\theta_m\}_{m=1}^{M}\right) = \pi\left(\theta_\ell \mid \{\theta_m\}_{m=1,m\neq\ell}^{M}\right) \pi\left(\{\theta_m\}_{m=1,m\neq\ell}^{M}\right)$$

The Gibbs sampler works by drawing random variates from the marginal densities $\pi\left(\theta_\ell \mid \{\theta_m\}_{m=1,m\neq\ell}^{M}\right)$ in a cyclic iterative pattern.

Gibbs Sampling

First iteration:

$$\theta_1^{(1)} \sim \pi\left(\theta_1 \mid \theta_2^{(0)}, \theta_3^{(0)}, \theta_4^{(0)}, \dots, \theta_M^{(0)}\right)$$

$$\theta_2^{(1)} \sim \pi\left(\theta_2 \mid \theta_1^{(1)}, \theta_3^{(0)}, \theta_4^{(0)}, \dots, \theta_M^{(0)}\right)$$

$$\theta_3^{(1)} \sim \pi\left(\theta_3 \mid \theta_1^{(1)}, \theta_2^{(1)}, \theta_4^{(0)}, \dots, \theta_M^{(0)}\right)$$

$$\vdots \qquad \vdots$$

$$\theta_M^{(1)} \sim \pi\left(\theta_M \mid \theta_1^{(1)}, \theta_2^{(1)}, \theta_4^{(1)}, \dots, \theta_{M-1}^{(1)}\right)$$

Gibbs Sampling

Second iteration:

$$\theta_1^{(2)} \sim \pi\left(\theta_1 \mid \theta_2^{(1)}, \theta_3^{(1)}, \theta_4^{(1)}, \dots, \theta_M^{(1)}\right)$$

$$\theta_2^{(2)} \sim \pi\left(\theta_2 \mid \theta_1^{(2)}, \theta_3^{(1)}, \theta_4^{(1)}, \dots, \theta_M^{(1)}\right)$$

$$\theta_3^{(2)} \sim \pi\left(\theta_3 \mid \theta_1^{(2)}, \theta_2^{(2)}, \theta_4^{(1)}, \dots, \theta_M^{(1)}\right)$$

$$\vdots \qquad \vdots$$

$$\theta_M^{(2)} \sim \pi\left(\theta_M \mid \theta_1^{(2)}, \theta_2^{(2)}, \theta_4^{(2)}, \dots, \theta_{M-1}^{(2)}\right)$$

Gibbs Sampling

$k+1$-th iteration:

$$\theta_1^{(k+1)} \sim \pi\left(\theta_1 \mid \theta_2^{(k)}, \theta_3^{(k)}, \theta_4^{(k)}, \dots, \theta_M^{(k)}\right)$$
$$\theta_2^{(k+1)} \sim \pi\left(\theta_2 \mid \theta_1^{(k+1)}, \theta_3^{(k)}, \theta_4^{(k)}, \dots, \theta_M^{(k)}\right)$$
$$\theta_3^{(k+1)} \sim \pi\left(\theta_3 \mid \theta_1^{(k+1)}, \theta_2^{(k+1)}, \theta_4^{(k)}, \dots, \theta_M^{(k)}\right)$$
$$\vdots \qquad \vdots$$
$$\theta_M^{(k+1)} \sim \pi\left(\theta_M \mid \theta_1^{(k)}, \theta_2^{(k)}, \theta_4^{(k)}, \dots, \theta_{M-1}^{(k)}\right)$$

At the end of the j-th iteration, the samples $\theta_0^{(j)}, \theta_1^{(j)}, \dots, \theta_M^{(j)}$ are considered to be drawn from the joint-density $\pi(\theta_0, \theta_1, \dots, \theta_M)$.

Stochastic Processes

Lecture Slideset 2
Stochastic Processes

Definition of a Stochastic Process

- Natural discrete-time signals can be characterised as random signals, since their values cannot be determined precisely; they are **unpredictable**.

Definition of a Stochastic Process

- Natural discrete-time signals can be characterised as random signals, since their values cannot be determined precisely; they are **unpredictable**.

- Consider an experiment with outcomes $\mathcal{S} = \{\zeta_k,\ k \in \mathbb{Z}^+\}$, each occurring with probability $\Pr(\zeta_k)$. Assign to each $\zeta_k \in \mathcal{S}$ a deterministic sequence $x[n, \zeta_k]$, $n \in \mathbb{Z}$.

Definition of a Stochastic Process

- Natural discrete-time signals can be characterised as random signals, since their values cannot be determined precisely; they are **unpredictable**.

- Consider an experiment with outcomes $\mathcal{S} = \{\zeta_k,\, k \in \mathbb{Z}^+\}$, each occurring with probability $\Pr(\zeta_k)$. Assign to each $\zeta_k \in \mathcal{S}$ a deterministic sequence $x[n, \zeta_k]$, $n \in \mathbb{Z}$.

- The sample space $\mathcal{S}$, probabilities $\Pr(\zeta_k)$, and the sequences $x[n, \zeta_k]$, $n \in \mathbb{Z}$ constitute a **discrete-time stochastic process**, or **random sequence**.

Definition of a Stochastic Process

- Natural discrete-time signals can be characterised as random signals, since their values cannot be determined precisely; they are **unpredictable**.

- Consider an experiment with outcomes $\mathcal{S} = \{\zeta_k,\ k \in \mathbb{Z}^+\}$, each occurring with probability $\Pr(\zeta_k)$. Assign to each $\zeta_k \in \mathcal{S}$ a deterministic sequence $x[n, \zeta_k]\ ,\ n \in \mathbb{Z}$.

- The sample space $\mathcal{S}$, probabilities $\Pr(\zeta_k)$, and the sequences $x[n, \zeta_k]\ ,\ n \in \mathbb{Z}$ constitute a **discrete-time stochastic process**, or **random sequence**.

- Formally, $x[n, \zeta_k]\ ,\ n \in \mathbb{Z}$ is a random sequence or **stochastic process** if, for a fixed value $n_0 \in \mathbb{Z}^+$ of n, $x[n_0, \zeta]\ ,\ n \in \mathbb{Z}$ is a random variable.

Definition of a Stochastic Process

- Natural discrete-time signals can be characterised as random signals, since their values cannot be determined precisely; they are **unpredictable**.

- Consider an experiment with outcomes $\mathcal{S} = \{\zeta_k,\, k \in \mathbb{Z}^+\}$, each occurring with probability $\Pr(\zeta_k)$. Assign to each $\zeta_k \in \mathcal{S}$ a deterministic sequence $x[n, \zeta_k]\,,\ n \in \mathbb{Z}$.

- The sample space $\mathcal{S}$, probabilities $\Pr(\zeta_k)$, and the sequences $x[n, \zeta_k]\,,\ n \in \mathbb{Z}$ constitute a **discrete-time stochastic process**, or **random sequence**.

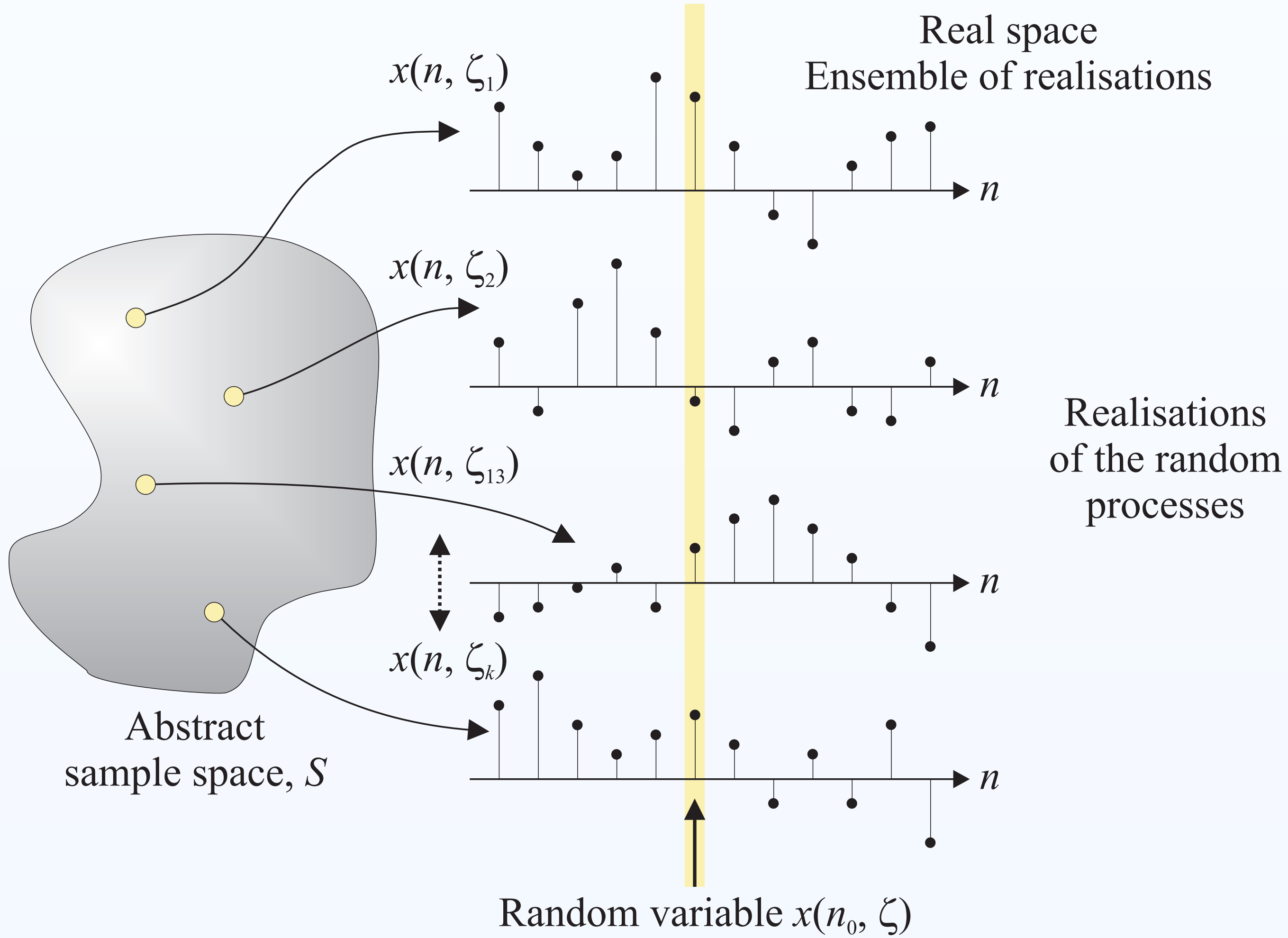
- Formally, $x[n, \zeta_k]\,,\ n \in \mathbb{Z}$ is a random sequence or **stochastic process** if, for a fixed value $n_0 \in \mathbb{Z}^+$ of n, $x[n_0, \zeta]\,,\ n \in \mathbb{Z}$ is a random variable.

- Also known as a **time series** in the statistics literature.

Definition of a Stochastic Process

- Natural discrete-time signals can be characterised as random signals, since their values cannot be determined precisely; they are **unpredictable**.
- Consider an experiment with outcomes $\mathcal{S} = \{\zeta_k,\, k \in \mathbb{Z}^+\}$, each occurring with probability $\Pr(\zeta_k)$. Assign to each $\zeta_k \in \mathcal{S}$ a deterministic sequence $x[n, \zeta_k]\,,\ n \in \mathbb{Z}$.
- The sample space $\mathcal{S}$, probabilities $\Pr(\zeta_k)$, and the sequences $x[n, \zeta_k]\,,\ n \in \mathbb{Z}$ constitute a **discrete-time stochastic process**, or **random sequence**.
- Formally, $x[n, \zeta_k]\,,\ n \in \mathbb{Z}$ is a random sequence or **stochastic process** if, for a fixed value $n_0 \in \mathbb{Z}^+$ of n, $x[n_0, \zeta]\,,\ n \in \mathbb{Z}$ is a random variable.
- Also known as a **time series** in the statistics literature.

Interpretation of Sequences



A graphical representation of a random process.

Interpretation of Sequences

Example. Consider a continuous-time random process, $x(t, \zeta)$, defined by a finite sized ensemble consisting of:

$$x(t,1) = -3\,u(t) \qquad x(t,2) = \cos(5\pi\,t)\;u(t)$$

$$x(t,3) = 10\,t\,u(t) \qquad x(t,4) = 2\sin(6\pi\,t + 0.2) \qquad \bowtie$$

1. Draw the ensemble.

2. For $t = 0.2$, determine the sample space.

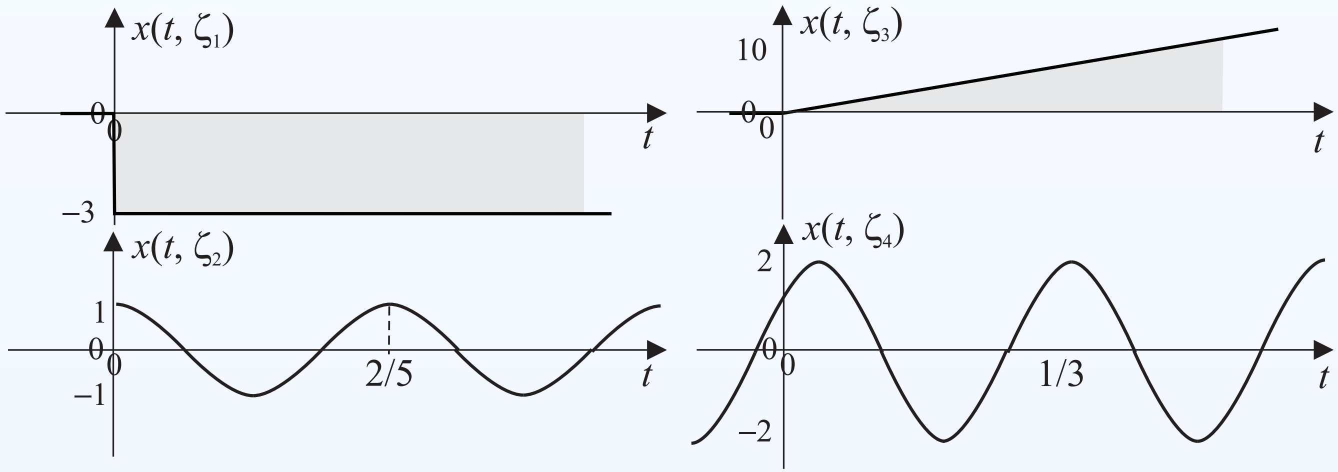
Interpretation of Sequences

Example. Consider a continuous-time random process, $x(t, \zeta)$, defined by a finite sized ensemble consisting of:

$$x(t,1) = -3\,u(t) \qquad x(t,2) = \cos(5\pi\,t)\;u(t)$$

$$x(t,3) = 10\,t\,u(t) \qquad x(t,4) = 2\sin(6\pi\,t + 0.2)$$

SOLUTION. 1. To plot the ensemble, draw all the realisations.



Ensemble of waveforms.

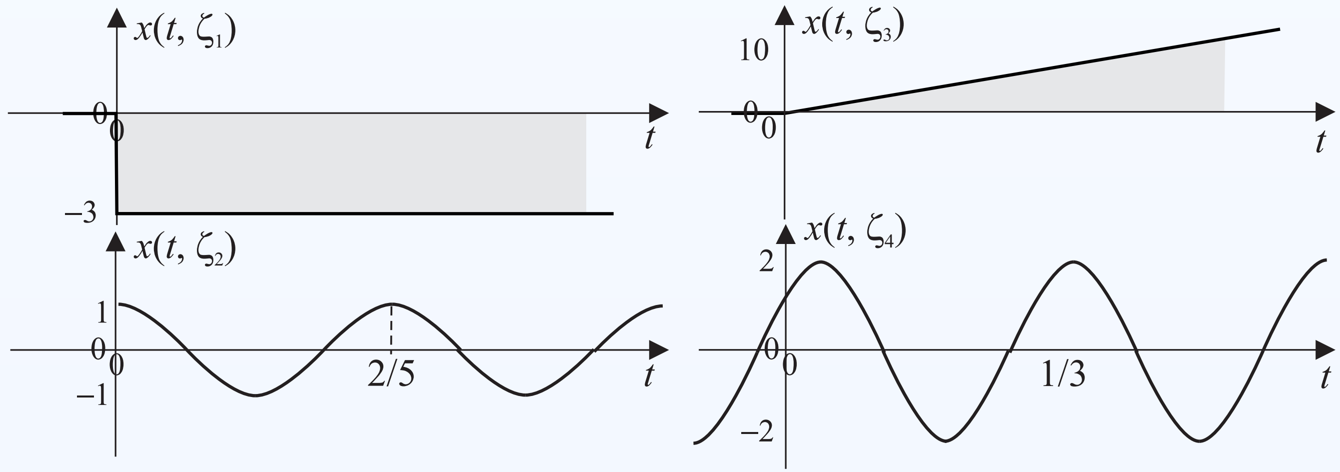
Interpretation of Sequences

Example. Consider a continuous-time random process, $x(t, \zeta)$, defined by a finite sized ensemble consisting of:

$$x(t,1) = -3\,u(t) \qquad x(t,2) = \cos(5\pi\,t)\;u(t)$$

$$x(t,3) = 10\,t\,u(t) \qquad x(t,4) = 2\sin(6\pi\,t + 0.2)$$

SOLUTION. 1. To plot the ensemble, draw all the realisations.



□

2. The sample space is thus $\{-3, -1, 2, -1.4736\}$.

Interpretation of Sequences

The set of all possible sequences $\{x[n, \zeta]\}$ is called an **ensemble**, and each individual sequence $x[n, \zeta_k]$, corresponding to a specific value of $\zeta = \zeta_k$, is called a **realisation** or a **sample sequence** of the ensemble.

Interpretation of Sequences

The set of all possible sequences $\{x[n, \zeta]\}$ is called an **ensemble**, and each individual sequence $x[n, \zeta_k]$, corresponding to a specific value of $\zeta = \zeta_k$, is called a **realisation** or a **sample sequence** of the ensemble.

There are four possible interpretations of $x[n, \zeta]$:

	ζ Fixed	ζ Variable
n Fixed	Number	Random variable
n Variable	Sample sequence	Stochastic process

Interpretation of Sequences

The set of all possible sequences $\{x[n, \zeta]\}$ is called an **ensemble**, and each individual sequence $x[n, \zeta_k]$, corresponding to a specific value of $\zeta = \zeta_k$, is called a **realisation** or a **sample sequence** of the ensemble.

There are four possible interpretations of $x[n, \zeta]$:

	ζ Fixed	ζ Variable
n Fixed	Number	Random variable
n Variable	Sample sequence	Stochastic process

Use simplified notation $x[n] \equiv x[n, \zeta]$ to denote both a stochastic process, and a single realisation. Use the terms **random process** and **stochastic process** interchangeably throughout this course.

Interpretation of Sequences

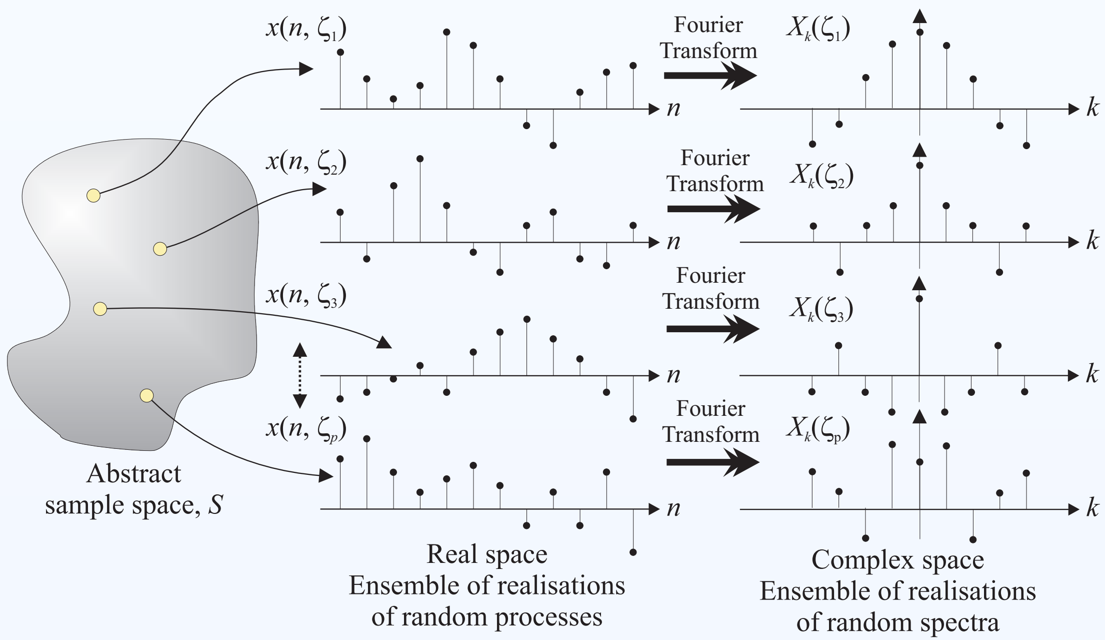
Building on these intepretations of sequences, this course will:

- The statistical properties of random signals, the statistical dependence of samples at different points in time.

Interpretation of Sequences

Building on these intepretations of sequences, this course will:

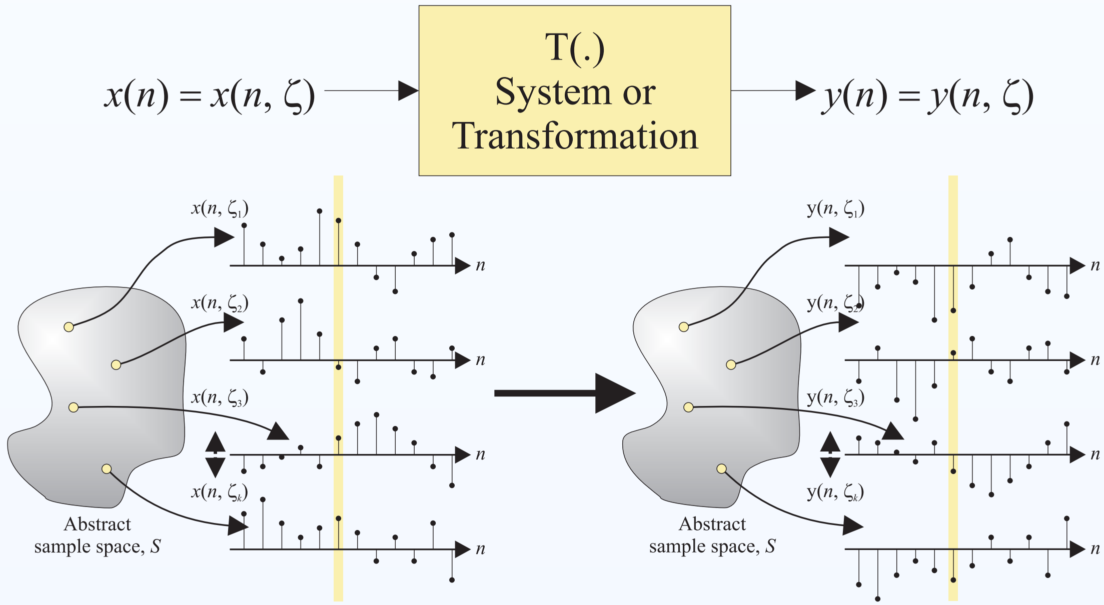
- The statistical properties of random signals, the statistical dependence of samples at different points in time.
- Interpreting stochastic signals in the frequency domain, the notion of a random spectrum, and the concept of the power spectral density.



Interpretation of Sequences

Building on these intepretations of sequences, this course will:

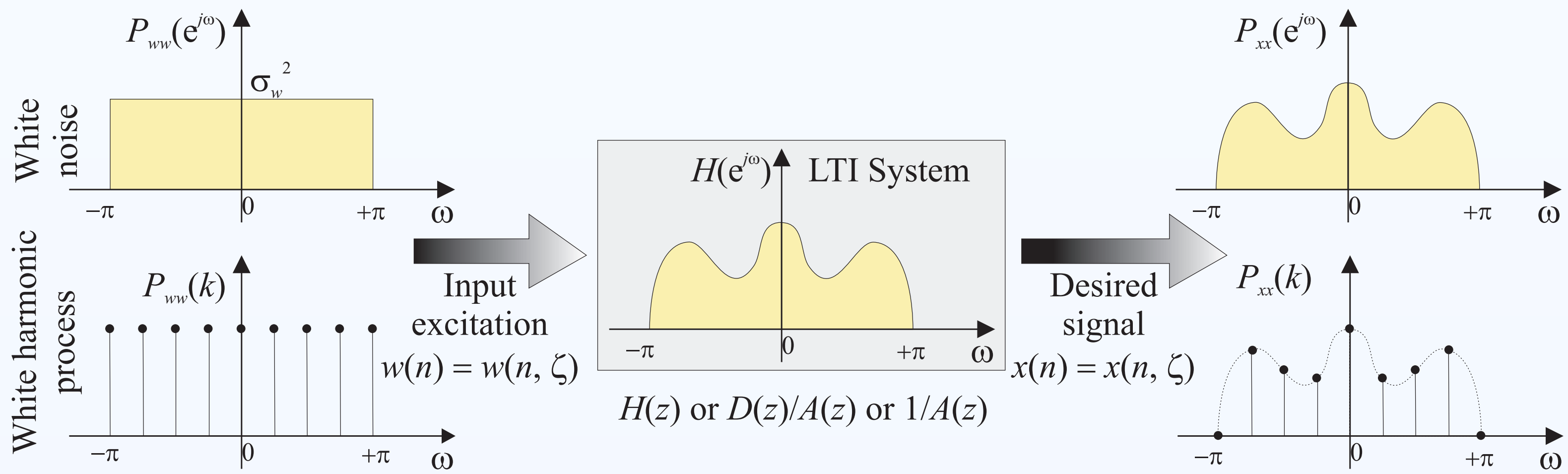
- What happens to a stochastic process and signals as it passes through systems?



Interpretation of Sequences

Building on these intepretations of sequences, this course will:

- What happens to a stochastic process and signals as it passes through systems?
- The notion of signal modelling for signal analysis and prediction.



Interpretation of Sequences

– End-of-Topic 45: **Introduction to the definition of stochastic processes** –

Any Questions?

Description using pdfs

For fixed $n = n_0$, $x[n_0, \zeta]$ is a random variable. Moreover, the random vector formed from the k random variables $\{x[n_j]\,,\, j \in \{1, \dots k\}\}$ is characterised by the cdf and pdfs:

$$F_X\left(x_1 \dots x_k \mid n_1 \dots n_k\right) = \Pr\left(x[n_1] \le x_1, \dots, x[n_k] \le x_k\right)$$

$$f_X\left(x_1 \dots x_k \mid n_1 \dots n_k\right) = \frac{\partial^k F_X\left(x_1 \dots x_k \mid n_1 \dots n_k\right)}{\partial x_1 \cdots \partial x_k}$$

Description using pdfs

For fixed $n = n_0$, $x[n_0, \zeta]$ is a random variable. Moreover, the random vector formed from the k random variables $\{x[n_j], j \in \{1, \dots k\}\}$ is characterised by the cdf and pdfs:

$$F_X(x_1 \dots x_k \mid n_1 \dots n_k) = \Pr(x[n_1] \le x_1, \dots, x[n_k] \le x_k)$$

$$f_X(x_1 \dots x_k \mid n_1 \dots n_k) = \frac{\partial^k F_X(x_1 \dots x_k \mid n_1 \dots n_k)}{\partial x_1 \cdots \partial x_k}$$

In exactly the same way as with random variables and random vectors, it is:

- difficult to estimate these probability functions without considerable additional information or assumptions;
- possible to frequently characterise stochastic processes usefully with much less information.

Second-order Statistical Description

Mean and Variance Sequence At time n, the **ensemble** mean and variance are given by:

$$\mu_x[n] = \mathbb{E}\left[x[n]\right]$$

$$\sigma_x^2[n] = \mathbb{E}\left[|x[n] - \mu_x[n]|^2\right] = \mathbb{E}\left[|x[n]|^2\right] - |\mu_x[n]|^2$$

Both $\mu_x[n]$ and $\sigma_x^2[n]$ are deterministic sequences.

Second-order Statistical Description

Mean and Variance Sequence At time n, the **ensemble** mean and variance are given by:

$$\mu_x[n] = \mathbb{E}\left[x[n]\right]$$

$$\sigma_x^2[n] = \mathbb{E}\left[|x[n] - \mu_x[n]|^2\right] = \mathbb{E}\left[|x[n]|^2\right] - |\mu_x[n]|^2$$

Both $\mu_x[n]$ and $\sigma_x^2[n]$ are deterministic sequences.

Autocorrelation sequence The second-order statistic $r_{xx}[n_1, n_2]$ provides a measure of the dependence between values of the process at two different times; it can provide information about the time variation of the process:

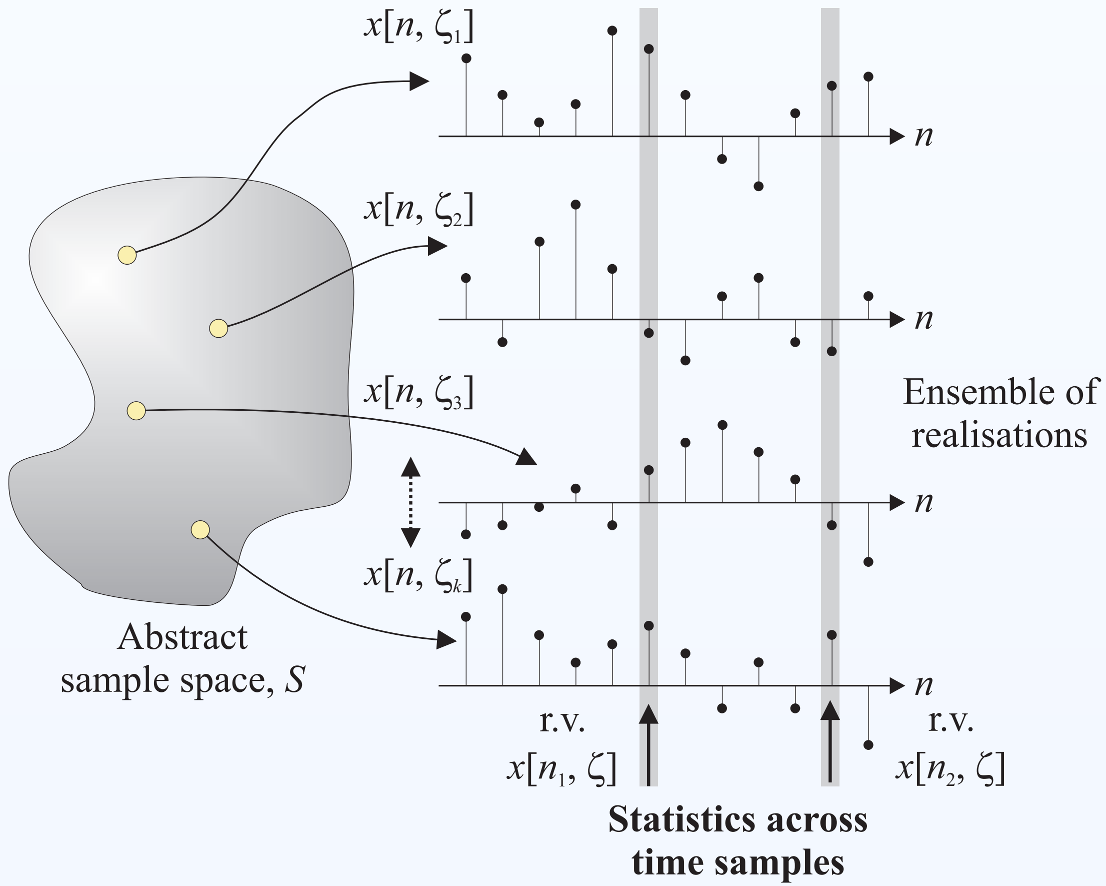
$$r_{xx}[n_1, n_2] = \mathbb{E}\left[x[n_1]\, x^*[n_2]\right]$$

- Note this definition is not consistent across all text book, or indeed University courses!

Second-order Statistical Description

Autocorrelation sequence provides a measure of the dependence between values of the process at two different times:

$$r_{xx}[n_1, n_2] = \mathbb{E}\left[x[n_1]\ x^*[n_2]\right]$$



Second-order Statistical Description

Autocovariance sequence The autocovariance sequence provides a measure of how similar the deviation from the mean of a process is at two different time instances:

$$\begin{aligned}\gamma_{xx}[n_1, n_2] &= \mathbb{E}\left[(x[n_1] - \mu_x[n_1])(x[n_2] - \mu_x[n_2])^*\right] \\ &= r_{xx}[n_1, n_2] - \mu_x[n_1]\ \mu_x^*[n_2]\end{aligned}$$

Second-order Statistical Description

Autocovariance sequence The autocovariance sequence provides a measure of how similar the deviation from the mean of a process is at two different time instances:

$$\begin{aligned}\gamma_{xx}[n_1, n_2] &= \mathbb{E}\left[(x[n_1] - \mu_x[n_1])(x[n_2] - \mu_x[n_2])^*\right] \\ &= r_{xx}[n_1, n_2] - \mu_x[n_1]\ \mu_x^*[n_2]\end{aligned}$$

To show how these deterministic sequences of a stochastic process can be calculated, several examples are considered in detail below.

Example of Calculating Autocorrelations

Example ([Manolakis:2000, Ex 3.9, page 144]). The harmonic process $x[n]$ is defined by:

$$x[n] = \sum_{k=1}^{M} A_k \cos(\omega_k n + \phi_k), \quad \omega_k \neq 0$$

where M, $\{A_k\}_1^M$ and $\{\omega_k\}_1^M$ are constants, and $\{\phi_k\}_1^M$ are pairwise independent random variables uniformly distributed in the interval $[0, 2\pi]$.

1. Determine the mean of $x[n]$.

2. Show the autocorrelation sequence is given by

$$r_{xx}[\ell] = \frac{1}{2} \sum_{k=1}^{M} |A_k|^2 \cos \omega_k \ell, \quad -\infty < \ell < \infty \qquad \bowtie$$

where $\ell \triangleq n_1 - n_2$, and $r_{xx}[\ell] \triangleq r_{xx}[n_1, n_1 + \ell]$ for any n_1.

Example of Calculating Autocorrelations

Example ([Manolakis:2000, Ex 3.9, page 144]). SOLUTION. 1. The expected value of the process is straightforwardly given by:

$$\mathbb{E}\left[x[n]\right] = \mathbb{E}\left[\sum_{k=1}^{M} A_k \cos(\omega_k n + \phi_k)\right] = \sum_{k=1}^{M} A_k \, \mathbb{E}\left[\cos(\omega_k n + \phi_k)\right]$$

Since a co-sinusoid is zero-mean, then:

$$\begin{aligned}\mathbb{E}\left[\cos(\omega_k n + \phi_k)\right] &= \int \cos(\omega_k n + \phi_k)\, f_{\Phi_k}(\phi_k)\, d\phi_k \\ &= \int_0^{2\pi} \cos(\omega_k n + \phi_k) \times \frac{1}{2\pi} \times d\phi_k = 0\end{aligned}$$

Hence, it follows:

$$\mathbb{E}\left[x[n]\right] = 0, \quad \forall n \qquad \square$$

Example of Calculating Autocorrelations

Example ([Manolakis:2000, Ex 3.9, page 144]). SOLUTION. 1. The autocorrelation $r_{xx}[n_1, n_2] = \mathbb{E}\left[x[n_1]\; x^*[n_2]\right]$ follows similarly:

$$r_{xx}[n_1, n_2] = \mathbb{E}\left[\sum_{k=1}^{M} A_k \cos(\omega_k n_1 + \phi_k) \sum_{j=1}^{M} A_j^* \cos(\omega_j n_2 + \phi_j)\right]$$

$$= \sum_{k=1}^{M} \sum_{j=1}^{M} A_k\, A_j^*\, \underbrace{\mathbb{E}\left[\cos(\omega_k n_1 + \phi_k)\, \cos(\omega_j n_2 + \phi_j)\right]}_{r(\phi_k, \phi_j)}$$

□

Example of Calculating Autocorrelations

Example ([Manolakis:2000, Ex 3.9, page 144]). SOLUTION. 1. The autocorrelation $r_{xx}[n_1, n_2] = \mathbb{E}\left[x[n_1]\; x^*[n_2]\right]$ follows similarly:

$$r_{xx}[n_1, n_2] = \mathbb{E}\left[\sum_{k=1}^{M} A_k \cos(\omega_k n_1 + \phi_k) \sum_{j=1}^{M} A_j^* \cos(\omega_j n_2 + \phi_j)\right]$$

$$= \sum_{k=1}^{M}\sum_{j=1}^{M} A_k\, A_j^*\, \underbrace{\mathbb{E}\left[\cos(\omega_k n_1 + \phi_k)\, \cos(\omega_j n_2 + \phi_j)\right]}_{r(\phi_k, \phi_j)}$$

After some algebra, it can be shown that:

$$\mathbb{E}\left[\cos(\omega_k n_1 + \phi_k)\, \cos(\omega_j n_2 + \phi_j)\right] = \begin{cases} \frac{1}{2}\cos\omega_k(n_1 - n_2) & k = j \\ 0 & \text{otherwise} \end{cases}$$

□

Example of Calculating Autocorrelations

Example ([Manolakis:2000, Ex 3.9, page 144]). SOLUTION. 1. After some algebra, it can be shown that:

$$\mathbb{E}\left[\cos(\omega_k n_1 + \phi_k)\, \cos(\omega_j n_2 + \phi_j)\right] = \begin{cases} \frac{1}{2}\cos\omega_k(n_1 - n_2) & k = j \\ 0 & \text{otherwise} \end{cases}$$

Substituting this expression into

$$r_{xx}[n_1, n_2] = \sum_{k=1}^{M}\sum_{j=1}^{M} A_k\, A_j^*\, \mathbb{E}\left[\cos(\omega_k n_1 + \phi_k)\, \cos(\omega_j n_2 + \phi_j)\right]$$

□

Example of Calculating Autocorrelations

Example ([Manolakis:2000, Ex 3.9, page 144]). SOLUTION. 1. After some algebra, it can be shown that:

$$\mathbb{E}\left[\cos(\omega_k n_1 + \phi_k)\,\cos(\omega_j n_2 + \phi_j)\right] = \begin{cases} \frac{1}{2}\cos\omega_k(n_1 - n_2) & k = j \\ 0 & \text{otherwise} \end{cases}$$

Substituting this expression into

$$r_{xx}[n_1, n_2] = \sum_{k=1}^{M}\sum_{j=1}^{M} A_k\, A_j^*\, \mathbb{E}\left[\cos(\omega_k n_1 + \phi_k)\,\cos(\omega_j n_2 + \phi_j)\right]$$

thus leads to the desired result, where $\ell = n_1 - n_2$:

$$r_{xx}[\ell] = \frac{1}{2}\sum_{k=1}^{M} |A_k|^2\,\cos\omega_k \ell, \quad -\infty < \ell < \infty \qquad \square$$

Example of Calculating Autocorrelations

Example (Functions of Random Process). A random variable $y[n]$ is defined to be:

$$y[n] = x[n] + x[n+m]$$

where m is some integer, and $x[n]$ is a stochastic process whose autocorrelation sequence (ACS) is given by:

$$r_{xx}[n_1, n_2] = e^{-(n_1-n_2)^2} \qquad \bowtie$$

Derive an expression for the ACS of the stochastic process $y[n]$, denoted $r_{yy}[n_1, n_2]$.

Example of Calculating Autocorrelations

Example (Functions of Random Process). A random variable $y[n]$ is

$$y[n] = x[n] + x[n+m]$$

where $x[n]$ is a stochastic process whose ACS is given by:

$$r_{xx}[n_1, n_2] = e^{-(n_1 - n_2)^2}$$

Derive an expression for $r_{yy}[n_1, n_2]$.

SOLUTION. In this example, it is simplest to form the product:

$$\begin{aligned} y[n_1]\, y^*[n_2] &= [x[n_1] + x[n_1+m]]\, [x^*[n_2] + x^*[n_2+m]] \\ &= x[n_1]\, x^*[n_2] + x[n_1]\, x^*[n_2+m] \\ &\quad + x[n_1+m]\, x^*[n_2] + x[n_1+m]\, x^*[n_2] \end{aligned}$$

□

Example of Calculating Autocorrelations

Example (Functions of Random Process). A random variable $y[n]$ is

$$y[n] = x[n] + x[n+m]$$

where $x[n]$ is a stochastic process whose ACS is given by:

$$r_{xx}[n_1, n_2] = e^{-(n_1 - n_2)^2}$$

Derive an expression for $r_{yy}[n_1, n_2]$.

SOLUTION. Then, taking expectations, it follows:

$$\begin{aligned} r_{yy}[n_1, n_2] &= r_{xx}[n_1, n_2] + r_{xx}[n_1, n_2 + m] \\ &\quad + r_{xx}[n_1 + m, n_2] + r_{xx}[n_1 + m, n_2 + m] \end{aligned}$$

□

Example of Calculating Autocorrelations

Example (Functions of Random Process). A random variable $y[n]$ is

$$y[n] = x[n] + x[n+m]$$

where $x[n]$ is a stochastic process whose ACS is given by:

$$r_{xx}[n_1, n_2] = e^{-(n_1-n_2)^2}$$

Derive an expression for $r_{yy}[n_1, n_2]$.

SOLUTION. Then, taking expectations, it follows:

$$r_{yy}[n_1, n_2] = r_{xx}[n_1, n_2] + r_{xx}[n_1, n_2+m] + r_{xx}[n_1+m, n_2] + r_{xx}[n_1+m, n_2+m] \quad \square$$

Using the result $r_{xx}[n_1, n_2] = e^{-(n_1-n_2)^2}$:

Example of Calculating Autocorrelations

Example (Functions of Random Process). A random variable $y[n]$ is

$$y[n] = x[n] + x[n+m]$$

where $x[n]$ is a stochastic process whose ACS is given by:

$$r_{xx}[n_1, n_2] = e^{-(n_1-n_2)^2}$$

Derive an expression for $r_{yy}[n_1, n_2]$.

SOLUTION. Using the result $r_{xx}[n_1, n_2] = e^{-(n_1-n_2)^2}$:

$$r_{yy}[r_1, r_2] = 2\, e^{-(n_1-n_2)^2} + e^{-(n_1-n_2+m)^2} + e^{-(n_1-n_2-m)^2} \qquad \square$$

Example of Calculating Autocorrelations

– End-of-Topic 46: **Statistical Description of a Stochastic Process** –

Any Questions?

Types of Stochastic Processes

Predictable Processes The unpredictability of a random process is, in general, the combined result of the following two characteristics:

1. The selection of a single realisation is based on the outcome of a random experiment;
2. No functional description is available for *all* realisations of the *ensemble*.

Types of Stochastic Processes

Predictable Processes The unpredictability of a random process is, in general, the combined result of the following two characteristics:

1. The selection of a single realisation is based on the outcome of a random experiment;
2. No functional description is available for *all* realisations of the *ensemble*.

In some special cases, however, a functional relationship is available. This means that after the occurrence of all samples of a particular realisation up to a particular point, n, all future values can be predicted exactly from the past ones.

Types of Stochastic Processes

Predictable Processes The unpredictability of a random process is, in general, the combined result of the following two characteristics:

1. The selection of a single realisation is based on the outcome of a random experiment;
2. No functional description is available for *all* realisations of the *ensemble*.

In some special cases, however, a functional relationship is available. This means that after the occurrence of all samples of a particular realisation up to a particular point, n, all future values can be predicted exactly from the past ones.

If this is the case for a random process, then it is called **predictable**, otherwise it is said to be **unpredictable** or a **regular process**.

Types of Stochastic Processes

Predictable Processes As an example of a predictable process, consider the signal:

$$x[n, \zeta] = A \sin(\omega\, n + \phi)$$

where A is a known amplitude, ω is a known normalised angular frequency, and ϕ is a random phase, where $\phi \sim f_\Phi(\phi)$ is its pdf.

Types of Stochastic Processes

Independence A stochastic process is independent iff

$$f_X\left(x_1, \dots, x_N \mid n_1, \dots, n_N\right) = \prod_{k=1}^{N} f_{X_k}\left(x_k \mid n_k\right)$$

$\forall N,\, n_k,\, k \in \{1, \dots, N\}$. Here, therefore, $x[n]$ is a sequence of independent random variables.

Types of Stochastic Processes

Independence A stochastic process is independent iff

$$f_X\left(x_1, \dots, x_N \mid n_1, \dots, n_N\right) = \prod_{k=1}^{N} f_{X_k}\left(x_k \mid n_k\right)$$

$\forall N,\, n_k,\, k \in \{1, \dots, N\}$. Here, therefore, $x[n]$ is a sequence of independent random variables.

An i. i. d. process is one where all the random variables $\{x[n_k, \zeta]\,,\, n_k \in \mathbb{Z}\}$ have the same pdf, and $x[n]$ will be called an **i. i. d.** random process.

Types of Stochastic Processes

Independence A stochastic process is independent iff

$$f_X\left(x_1, \dots, x_N \mid n_1, \dots, n_N\right) = \prod_{k=1}^{N} f_{X_k}\left(x_k \mid n_k\right)$$

$\forall N,\, n_k,\, k \in \{1, \dots, N\}$. Here, therefore, $x[n]$ is a sequence of independent random variables.

An i. i. d. process is one where all the random variables $\{x[n_k, \zeta]\,,\, n_k \in \mathbb{Z}\}$ have the same pdf, and $x[n]$ will be called an **i. i. d.** random process.

An uncorrelated processes is a sequence of uncorrelated random variables:

$$\gamma_{xx}[n_1, n_2] = \sigma_x^2[n_1]\; \delta[n_1 - n_2]$$

Types of Stochastic Processes

An orthogonal process is a sequence of **orthogonal random variables**, and is given by:

$$r_{xx}[n_1, n_2] = \mathbb{E}\left[|x[n_1]|^2\right] \delta[n_1 - n_2]$$

If a process is zero-mean, then it is both **orthogonal** and **uncorrelated** since $\gamma_{xx}[n_1, n_2] = r_{xx}[n_1, n_2]$.

Types of Stochastic Processes

An orthogonal process is a sequence of **orthogonal random variables**, and is given by:

$$r_{xx}[n_1, n_2] = \mathbb{E}\left[|x[n_1]|^2\right] \delta[n_1 - n_2]$$

If a process is zero-mean, then it is both **orthogonal** and **uncorrelated** since $\gamma_{xx}[n_1, n_2] = r_{xx}[n_1, n_2]$.

A stationary process is a random process where its statistical properties do not vary with time. Processes whose statistical properties **do** change with time are referred to as **nonstationary**.

Types of Stochastic Processes

– End-of-Topic 47: **Types of Random Signals** –

Any Questions?

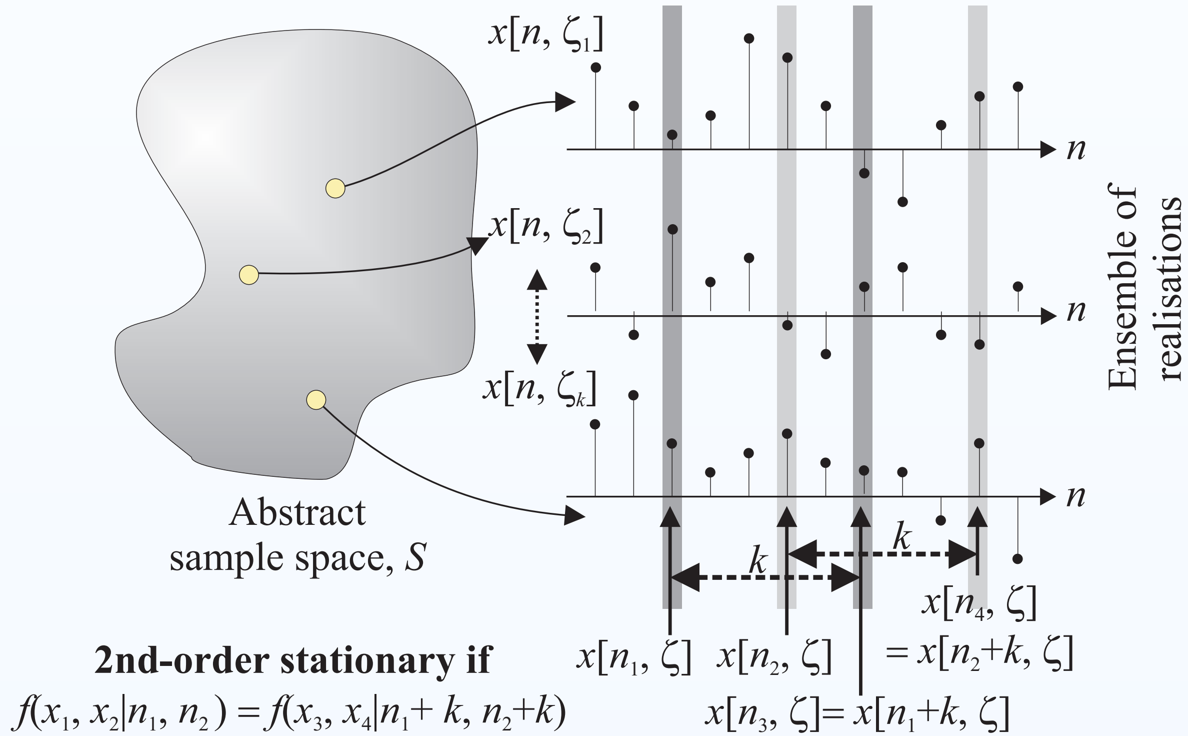
Stationary Processes

A random process $x[n]$ has been called **stationary** if its statistics determined for $x[n]$ are equal to those for $x[n+k]$, for every k. There are various formal definitions of **stationarity**, along with **quasi-stationary** processes, which are discussed below.

- **Order-N and strict-sense stationarity**
- **Wide-sense stationarity**
- Autocorrelation properties for WSS processes
- **Wide-sense periodicity** and **cyclo-stationarity**
- Local- or **quasi-stationary** processes

After this, some examples of various stationary processes will be given.

Order-N and strict-sense stationarity

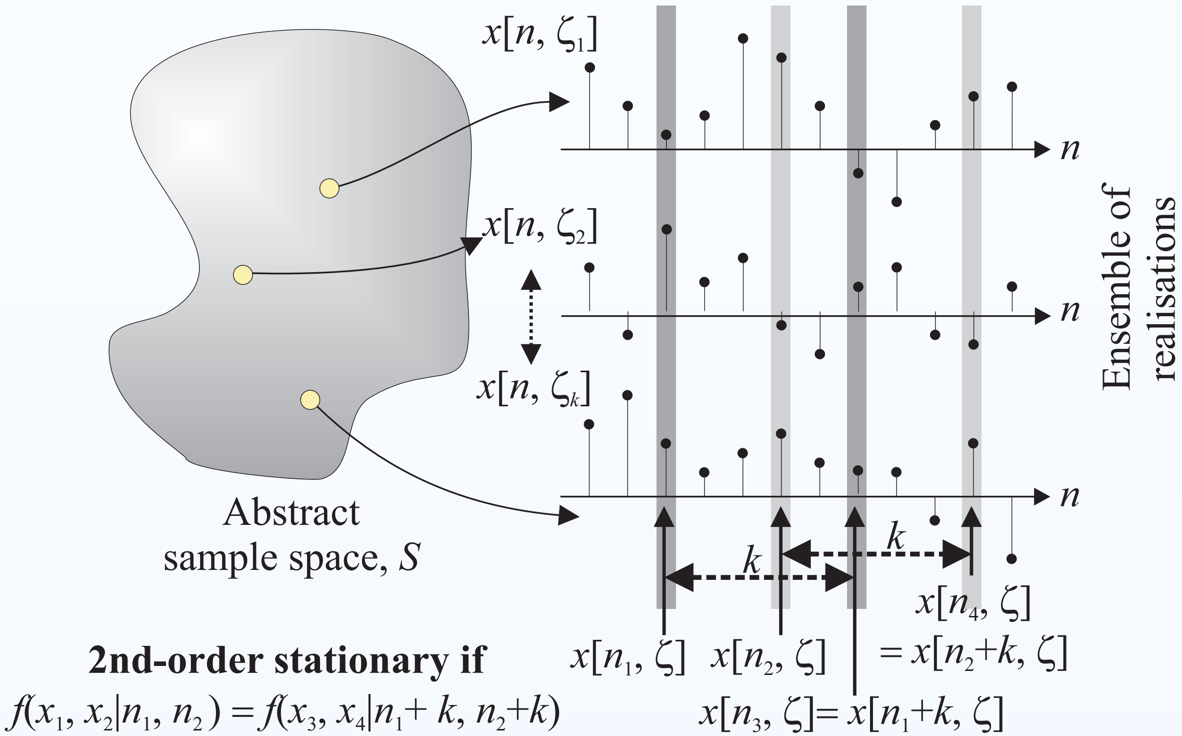


Definition (Stationary of order-N). A stochastic process $x[n]$ is called **stationary of order-N** if for any value of k then:

$$f_X(x_1, \dots, x_N \mid n_1, \dots, n_N) = f_X(x_1, \dots, x_N \mid n_1 + k, \dots, n_N + k)$$

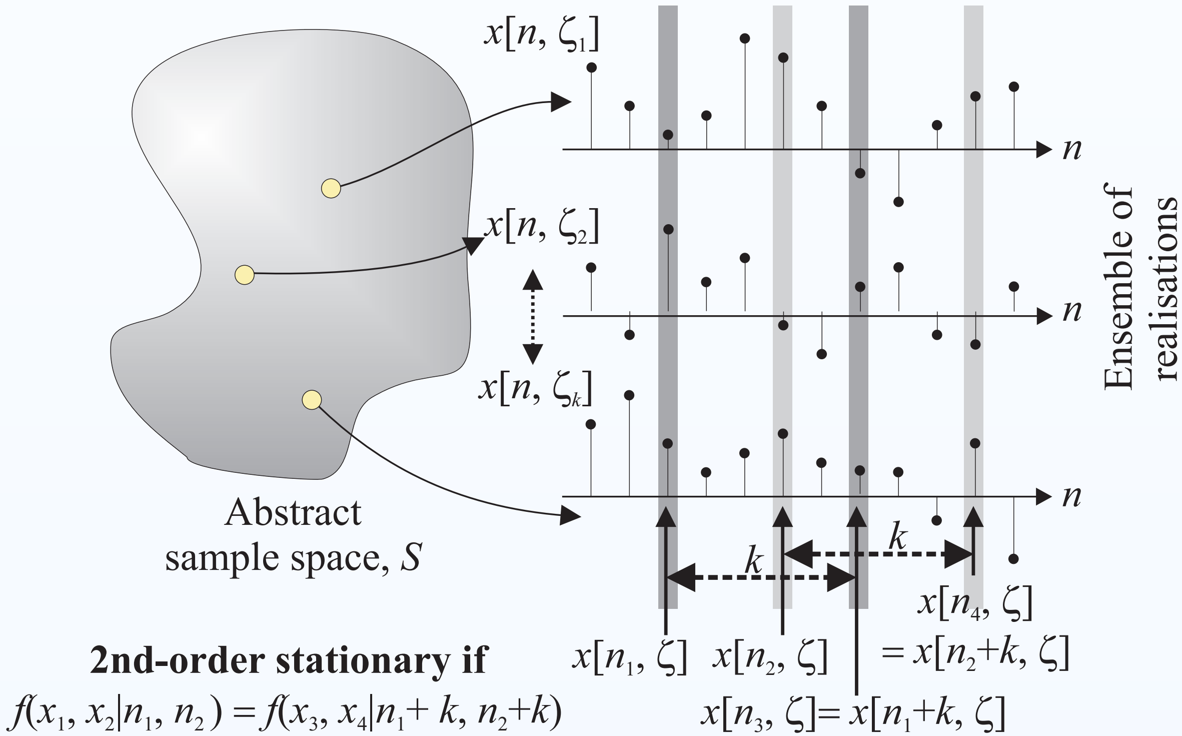
♢

Order-N and strict-sense stationarity



Definition (Strict-sense stationary). If $x[n]$ is stationary for all orders $N \in \mathbb{Z}^+$, it is said to be **strict-sense stationary (SSS)**.

Order-N and strict-sense stationarity



Definition (Strict-sense stationary). If $x[n]$ is stationary for all orders $N \in \mathbb{Z}^+$, it is said to be **SSS**.

- However, SSS is more restrictive than necessary in practical applications, and is a rarely required property.

Wide-sense stationarity

A more relaxed form of stationarity, which is sufficient for practical problems, occurs when a random process is stationary order-2; such a process is **wide-sense stationary (WSS)**.

Wide-sense stationarity

Definition (Wide-sense stationarity). A random signal $x[n]$ is called wide-sense stationary if:

- the mean and variance is constant and independent of n:

$$\mathbb{E}\left[x[n]\right] = \mu_x$$
$$\operatorname{var}\left[x[n]\right] = \sigma_x^2$$

- the autocorrelation depends only on the time difference $\ell = n_1 - n_2$, called the lag:

$$\begin{aligned} r_{xx}[n_1, n_2] &= r_{xx}^*[n_2, n_1] = \mathbb{E}\left[x[n_1]\; x^*[n_2]\right] \\ &= r_{xx}[\ell] = r_{xx}[n_1 - n_2] = \mathbb{E}\left[x[n_1]\; x^*[n_1 - \ell]\right] \qquad \diamond \\ &= \mathbb{E}\left[x[n_2 + \ell]\; x^*[n_2]\right] \end{aligned}$$

Wide-sense stationarity

- The definition of the **lag** is not consistent across textbooks, or indeed courses on this MSc!
- Elsewhere, the following definition is used:

$$r_{xx}[n_1, n_2] \triangleq \mathbb{E}\left[x[n_1]\; x^*\left[n_1 + \hat{\ell}\right]\right]$$

$$r_{xx}\left[\hat{\ell}\right] \triangleq \mathbb{E}\left[x\left[n - \hat{\ell}\right]\; x^*[n]\right]$$

Wide-sense stationarity

- The definition of the **lag** is not consistent across textbooks, or indeed courses on this MSc!
- Elsewhere, the following definition is used:

$$r_{xx}[n_1, n_2] \triangleq \mathbb{E}\left[x[n_1]\, x^*\left[n_1 + \hat{\ell}\right]\right]$$

$$r_{xx}\left[\hat{\ell}\right] \triangleq \mathbb{E}\left[x\left[n - \hat{\ell}\right]\, x^*[n]\right]$$

- Although a minor change in sign, this does have implications when considering results that are functions of random processes, such as a signal passing through a linear system, or frequency-domain analysis.

Wide-sense stationarity

- The definition of the **lag** is not consistent across textbooks, or indeed courses on this MSc!
- Elsewhere, the following definition is used:

$$r_{xx}[n_1, n_2] \triangleq \mathbb{E}\left[x[n_1]\; x^*\left[n_1 + \hat{\ell}\right]\right]$$

$$r_{xx}\left[\hat{\ell}\right] \triangleq \mathbb{E}\left[x\left[n - \hat{\ell}\right]\; x^*[n]\right]$$

- Although a minor change in sign, this does have implications when considering results that are functions of random processes, such as a signal passing through a linear system, or frequency-domain analysis.
- It is simply something to become used to, and to understand the equations and use the appropriate subsequent results carefully.

Wide-sense stationarity

- The autocovariance sequence is given by:

$$\gamma_{xx}[\ell] = r_{xx}[\ell] - |\mu_x|^2$$

Wide-sense stationarity

- The autocovariance sequence is given by:

$$\gamma_{xx}[\ell] = r_{xx}[\ell] - |\mu_x|^2$$

- Since 2nd-order moments are defined in terms of 2nd-order pdf, then strict-sense stationary are always WSS, but not necessarily *vice-versa*, except if the signal is Gaussian.

Wide-sense stationarity

- The autocovariance sequence is given by:

$$\gamma_{xx}[\ell] = r_{xx}[\ell] - |\mu_x|^2$$

- Since 2nd-order moments are defined in terms of 2nd-order pdf, then strict-sense stationary are always WSS, but not necessarily *vice-versa*, except if the signal is Gaussian.

- In practice, however, it is very rare to encounter a signal that is stationary in the wide-sense, but not stationary in the strict sense.

Wide-sense stationarity

Example (Sum of sinusoids). A discrete-time random process, $g[n]$, is defined as

$$g[n] = A \sin(\omega_0 n) + B \cos(\omega_0 n) \qquad \bowtie$$

where A and B are independent random variables each having zero mean and variance σ^2, ω_0 is a fixed frequency, and n is the time-index.

- Determine the mean and autocovariance function of $g[n]$.
- Determine whether or not $g[n]$ is a WSS process. Explain your answer.

Wide-sense stationarity

Example (Sum of sinusoids). A process, $g[n]$, is defined as

$$g[n] = A \sin(\omega_0 n) + B \cos(\omega_0 n)$$

where A and B are independent random variables each having zero mean and variance σ^2, ω_0 is a fixed frequency.

- Determine the mean and autocovariance function of $g[n]$.
- Determine whether or not $g[n]$ is a WSS process.

SOLUTION.

- Noting that the expectation operator is linear:

$$\mu_g[n] = \mathbb{E}[g[n]] = \mathbb{E}[A \sin \omega_0 n] + \mathbb{E}[B \cos \omega_0 n]$$

$\sin \omega_0 n$ and $\cos \omega_0 n$ are deterministic & $\mathbb{E}[A] = \mathbb{E}[B] = 0$:

$$\mu_g[n] = \mathbb{E}[A] \sin \omega_0 n + \mathbb{E}[B] \cos \omega_0 n = 0$$

Wide-sense stationarity

Example (Sum of sinusoids). A process, $g[n]$, is defined as

$$g[n] = A \sin(\omega_0 n) + B \cos(\omega_0 n)$$

where A and B are independent random variables each having zero mean and variance σ^2, ω_0 is a fixed frequency.

- Determine whether or not $g[n]$ is a WSS process.

SOLUTION. - The autocovariance function is given by:

$$\gamma_{gg}[n_1, n_2] = \mathbb{E}\left[(g[n_1] - \mu_g[n_1])(g[n_2] - \mu_g[n_2])\right]$$

□

Wide-sense stationarity

Example (Sum of sinusoids). A process, $g[n]$, is defined as

$$g[n] = A \sin(\omega_0 n) + B \cos(\omega_0 n)$$

where A and B are independent random variables each having zero mean and variance σ^2, ω_0 is a fixed frequency.

- Determine whether or not $g[n]$ is a WSS process.

SOLUTION. - The autocovariance function is given by:

$$\gamma_{gg}[n_1, n_2] = \mathbb{E}\left[(g[n_1] - \mu_g[n_1])(g[n_2] - \mu_g[n_2])\right] \qquad \square$$

Hence, since $\mu_g[n_i] = 0$, it follows:

Wide-sense stationarity

Example (Sum of sinusoids). A process, $g[n]$, is defined as

$$g[n] = A \sin(\omega_0 n) + B \cos(\omega_0 n)$$

where A and B are independent random variables each having zero mean and variance σ^2, ω_0 is a fixed frequency.

- Determine whether or not $g[n]$ is a WSS process.

SOLUTION. ● The autocovariance function is given by:

$$\begin{aligned}\gamma_{gg}[n_1, n_2] &= \mathbb{E}\left[(A\sin\omega_0 n_1 + B\cos\omega_0 n_1)(A\sin\omega_0 n_2 + B\cos\omega_0 n_2)\right] \\ &= \mathbb{E}\left[A^2\right]\sin\omega_0 n_1\ \sin\omega_0 n_2 + \mathbb{E}\left[AB\right]\sin\omega_0 n_1\ \cos\omega_0 n_2 \\ &\quad + \mathbb{E}\left[BA\right]\cos\omega_0 n_1\ \sin\omega_0 n_2 + \mathbb{E}\left[B^2\right]\cos\omega_0 n_1\ \cos\omega_0 n_2\end{aligned}$$

□

Wide-sense stationarity

Example (Sum of sinusoids). A process, $g[n]$, is defined as

$$g[n] = A\,\sin\left(\omega_0 n\right) + B\,\cos\left(\omega_0 n\right)$$

where A and B are independent random variables each having zero mean and variance σ^2, ω_0 is a fixed frequency.

- Determine whether or not $g[n]$ is a WSS process.

SOLUTION. The autocovariance function is given by:

$$\begin{aligned}\gamma_{gg}[n_1, n_2] &= \mathbb{E}\left[\left(A\sin\omega_0 n_1 + B\cos\omega_0 n_1\right)\left(A\sin\omega_0 n_2 + B\cos\omega_0 n_2\right)\right] \\ &= \mathbb{E}\left[A^2\right]\sin\omega_0 n_1\,\sin\omega_0 n_2 + \mathbb{E}\left[AB\right]\sin\omega_0 n_1\,\cos\omega_0 n_2 \\ &\quad + \mathbb{E}\left[BA\right]\cos\omega_0 n_1\,\sin\omega_0 n_2 + \mathbb{E}\left[B^2\right]\cos\omega_0 n_1\,\cos\omega_0 n_2 \square\end{aligned}$$

$A\&B$ are independent, $\mathbb{E}\left[AB\right] = \mathbb{E}\left[BA\right] = \mathbb{E}\left[A\right]\mathbb{E}\left[B\right] = 0$.

Wide-sense stationarity

Example (Sum of sinusoids). A process, $g[n]$, is defined as

$$g[n] = A\,\sin\left(\omega_0 n\right) + B\,\cos\left(\omega_0 n\right)$$

where A and B are independent random variables each having zero mean and variance σ^2, ω_0 is a fixed frequency.

- Determine whether or not $g[n]$ is a WSS process.

SOLUTION. - Noting $\text{var}\left[A\right] = \text{var}\left[B\right] = \sigma^2$ and

$$\text{var}\left[A\right] = \mathbb{E}\left[A^2\right] - \mathbb{E}^2\left[A\right] \qquad \square$$

means that $\mathbb{E}\left[A^2\right] = \mathbb{E}\left[B^2\right] = \sigma^2$.

Wide-sense stationarity

Example (Sum of sinusoids). A process, $g[n]$, is defined as

$$g[n] = A \sin(\omega_0 n) + B \cos(\omega_0 n)$$

where A and B are independent random variables each having zero mean and variance σ^2, ω_0 is a fixed frequency.

- Determine whether or not $g[n]$ is a WSS process.

SOLUTION. - Noting $\text{var}[A] = \text{var}[B] = \sigma^2$ and

$$\text{var}[A] = \mathbb{E}[A^2] - \mathbb{E}^2[A]$$

means that $\mathbb{E}[A^2] = \mathbb{E}[B^2] = \sigma^2$. Thus,

$$\gamma_{gg}[n_1, n_2] = \sigma^2 (\sin\omega_0 n_1 \sin\omega_0 n_2 + \cos\omega_0 n_1 \cos\omega_0 n_2)$$

□

Wide-sense stationarity

Example (Sum of sinusoids). A process, $g[n]$, is defined as

$$g[n] = A \sin(\omega_0 n) + B \cos(\omega_0 n)$$

where A and B are independent random variables each having zero mean and variance σ^2, ω_0 is a fixed frequency.

- Determine whether or not $g[n]$ is a WSS process.

SOLUTION. Thus,

$$\gamma_{gg}[n_1, n_2] = \sigma^2 (\sin \omega_0 n_1 \sin \omega_0 n_2 + \cos \omega_0 n_1 \cos \omega_0 n_2)$$

Using the supplied trigonometric identity, it follows that:

$$\gamma_{gg}[n_1, n_2] = \sigma^2 \cos \omega_0 (n_1 - n_2) \qquad \square$$

Wide-sense stationarity

Example (Sum of sinusoids). A process, $g[n]$, is defined as

$$g[n] = A \sin(\omega_0 n) + B \cos(\omega_0 n)$$

where A and B are independent random variables each having zero mean and variance σ^2, ω_0 is a fixed frequency.

- Determine whether or not $g[n]$ is a WSS process.

SOLUTION. - To be WSS, the mean and variance must be constant, and the ACS a function of $n_1 - n_2$. The ACS is:

$$\begin{aligned} r_{gg}[n_1, n_2] &= \gamma_{gg}[n_1, n_2] + \mu_g[n_1]\,\mu_g[n_2] \\ &= \sigma^2 \cos\omega_0 (n_1 - n_2) \end{aligned} \qquad \square$$

Thus, mean is constant, and the ACS is a function of the time difference $n_1 - n_2$ only. Therefore it is WSS.

Wide-sense stationarity

– End-of-Topic 48: **Overview of types of stationary processes, and examples of WSS processes** –

Any Questions?

WSS Properties

The average power of a WSS process $x[n]$ satisfies:

$$r_{xx}[0] = \sigma_x^2 + |\mu_x|^2 \geq 0$$

$$r_{xx}[0] \geq |r_{xx}[\ell]|, \quad \text{for all } \ell$$

WSS Properties

The average power of a WSS process $x[n]$ satisfies:

$$r_{xx}[0] = \sigma_x^2 + |\mu_x|^2 \geq 0$$
$$r_{xx}[0] \geq |r_{xx}[\ell]|, \quad \text{for all } \ell$$

The expression for power can be broken down as follows:

Average DC Power: $|\mu_x|^2$

Average AC Power: σ_x^2

Total average power: $r_{xx}[0] \geq 0$

$$\text{Total average power} = \text{Average DC power} + \text{Average AC power}$$

WSS Properties

The average power of a WSS process $x[n]$ satisfies:

$$r_{xx}[0] = \sigma_x^2 + |\mu_x|^2 \geq 0$$
$$r_{xx}[0] \geq |r_{xx}[\ell]|, \quad \text{for all } \ell$$

The expression for power can be broken down as follows:

Average DC Power: $|\mu_x|^2$

Average AC Power: σ_x^2

Total average power: $r_{xx}[0] \geq 0$

$$\text{Total average power} = \text{Average DC power} + \text{Average AC power}$$

Moreover, it follows that $\gamma_{xx}[0] \geq |\gamma_{xx}[\ell]|$.

WSS Properties

The ACS $r_{xx}[\ell]$ satisfies two more properties:

- a conjugate symmetric function of the lag ℓ:

$$r_{xx}^*[-\ell] = r_{xx}[\ell]$$

- a **nonnegative-definite** or **positive semi-definite** function, such that for any sequence $\alpha[n]$:

$$\sum_{n=1}^{M} \sum_{m=1}^{M} \alpha^*[n]\; r_{xx}[n-m]\; \alpha[m] \geq 0$$

WSS Properties

The ACS $r_{xx}[\ell]$ satisfies two more properties:

- a conjugate symmetric function of the lag ℓ:

$$r_{xx}^*[-\ell] = r_{xx}[\ell]$$

- a **nonnegative-definite** or **positive semi-definite** function, such that for any sequence $\alpha[n]$:

$$\sum_{n=1}^{M} \sum_{m=1}^{M} \alpha^*[n] \; r_{xx}[n-m] \; \alpha[m] \geq 0$$

Note that, more generally, even a correlation function for a nonstationary random process is **positive semi-definite**:

$$\sum_{n=1}^{M} \sum_{m=1}^{M} \alpha^*[n] \; r_{xx}[n,m] \, \alpha[m] \geq 0 \quad \text{for any sequence } \alpha[n]$$

WSS Properties

Example (Cosinusoid). The function $r[\ell] = \cos \omega_0 \ell$ is claimed to be a valid ACS. Test the properties of this function to determine if this is claim is true or not.

WSS Properties

Example (Cosinusoid). The function $r[\ell] = \cos \omega_0 \ell$ is claimed to be a valid ACS. Test the properties of this function to determine if this is claim is true or not.

SOLUTION. The function $r[\ell] = \cos \omega_0 \ell$ satisfies:

- the symmetric property, $r[\ell] = r[-\ell]$;
- the equality $r[0] \geq |r[\ell]|$ for all ℓ;
- and $r[0] \geq 0$ □

.

WSS Properties

Example (Cosinusoid). The function $r[\ell] = \cos \omega_0 \ell$ is claimed to be a valid ACS. Test the properties of this function to determine if this is claim is true or not.

SOLUTION. The final property of positive semi-definiteness is a little more tedious to verify. Let:

$$I = \sum_{n=1}^{M} \sum_{m=1}^{M} \alpha^*[n] \; r_{xx}[n-m] \; \alpha[m]$$

$$= \sum_{n=1}^{M} \sum_{m=1}^{M} \alpha[n] \, \alpha[m] \; \cos \omega_0 \, (n-m)$$

□

WSS Properties

Example (Cosinusoid). The function $r[\ell] = \cos \omega_0 \ell$ is claimed to be a valid ACS. Test the properties of this function to determine if this is claim is true or not.

SOLUTION. The final property of positive semi-definiteness is a little more tedious to verify. Let:

$$I = \sum_{n=1}^{M} \sum_{m=1}^{M} \alpha^*[n]\, r_{xx}[n-m]\, \alpha[m]$$

$$= \sum_{n=1}^{M} \sum_{m=1}^{M} \alpha[n]\, \alpha[m]\, \cos \omega_0 (n-m) \qquad \square$$

Using the trigonometric identity:
$\cos \omega_0 (n-m) = \cos \omega_0 n \, \cos \omega_0 m + \sin \omega_0 n \, \sin \omega_0 m$, then consider the resulting first term and using the fact $r[\ell]$ is real:

WSS Properties

Example (Cosinusoid). The function $r[\ell] = \cos \omega_0 \ell$ is claimed to be a valid ACS. Test the properties of this function to determine if this is claim is true or not.

SOLUTION. Using the trigonometric identity: $\cos \omega_0 (n-m) = \cos \omega_0 n \, \cos \omega_0 m + \sin \omega_0 n \, \sin \omega_0 m$, then consider the resulting first term and using the fact $r[\ell]$ is real:

$$
\begin{aligned}
I_1 &= \sum_{n=1}^{M} \sum_{m=1}^{M} \alpha[n]\, \alpha[m] \, \cos \omega_0 n \, \cos \omega_0 m \\
&= \left(\sum_{n=1}^{M} \alpha[n] \, \cos \omega_0 n \right) \left(\sum_{m=1}^{M} \alpha[m] \cos \omega_0 m \right) \\
&= \left(\sum_{n=1}^{M} \alpha[n] \, \cos \omega_0 n \right)^2 \geq 0
\end{aligned}
$$

WSS Properties

Example (Cosinusoid). The function $r[\ell] = \cos \omega_0 \ell$ is claimed to be a valid ACS. Test the properties of this function to determine if this is claim is true or not.

SOLUTION. Using the trigonometric identity: :

$$I_1 = \sum_{n=1}^{M} \sum_{m=1}^{M} \alpha[n]\, \alpha[m]\, \cos \omega_0 n\, \cos \omega_0 m$$

$$= \left(\sum_{n=1}^{M} \alpha[n]\, \cos \omega_0 n \right) \left(\sum_{m=1}^{M} \alpha[m] \cos \omega_0 m \right)$$

$$= \left(\sum_{n=1}^{M} \alpha[n]\, \cos \omega_0 n \right)^2 \geq 0 \qquad \square$$

A similar argument can be made for the second term, $\Rightarrow I \geq 0$.

WSS Properties

- The Fourier transform of an autocorrelation sequence (ACS) or autocorrelation function (ACF) is an extremely important concept, called the power spectral density (PSD) which will be discussed in the next handout.

WSS Properties

- The Fourier transform of an autocorrelation sequence (ACS) or autocorrelation function (ACF) is an extremely important concept, called the PSD which will be discussed in the next handout.
- It will be proved that the PSD should always be positive.

WSS Properties

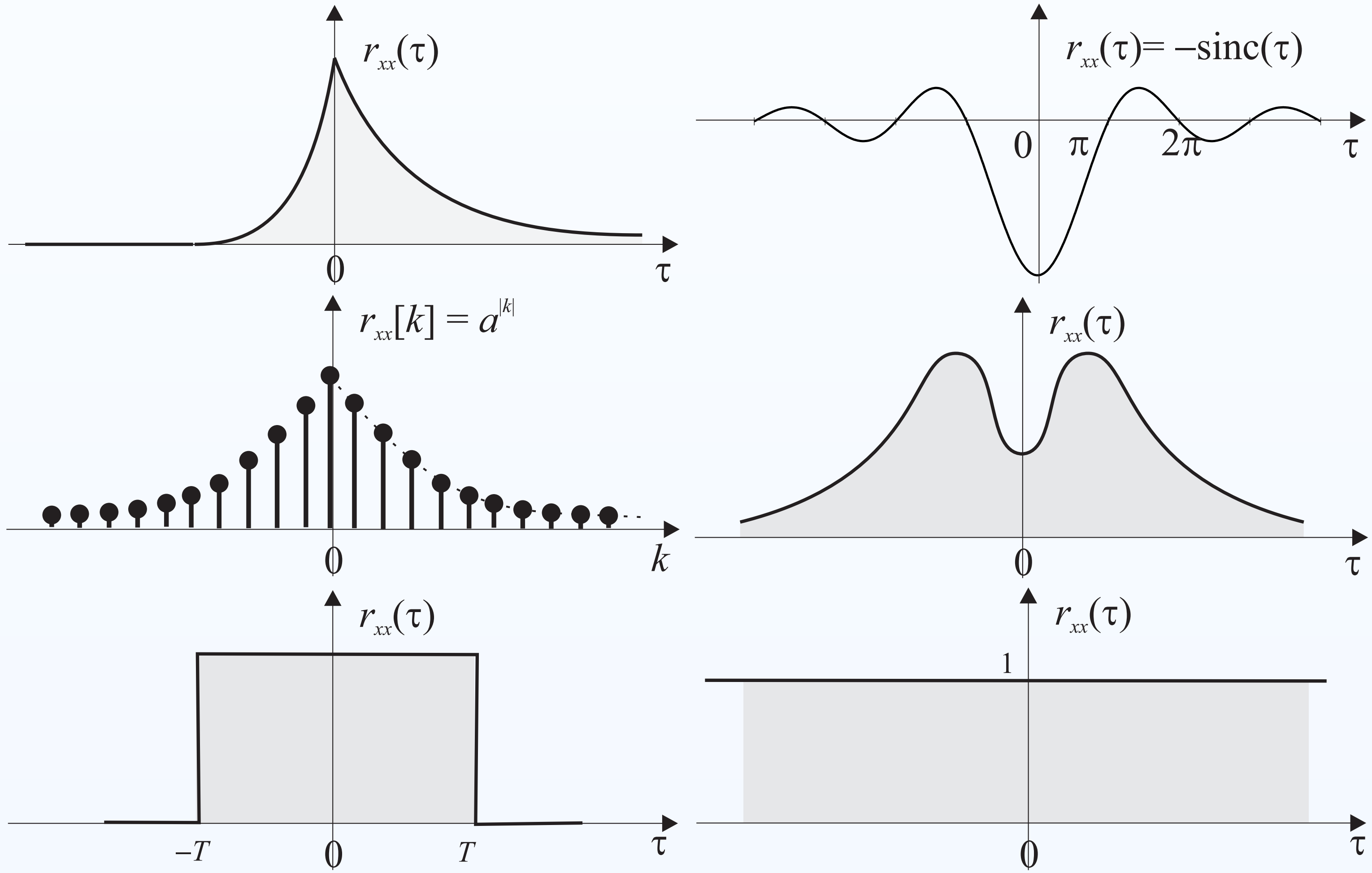
- The Fourier transform of an autocorrelation sequence (ACS) or autocorrelation function (ACF) is an extremely important concept, called the PSD which will be discussed in the next handout.
- It will be proved that the PSD should always be positive.
- It is easy to prove that an ACS or ACF has a positive Fourier transform if, and only if, it is positive semi-definite.

WSS Properties

- The Fourier transform of an autocorrelation sequence (ACS) or autocorrelation function (ACF) is an extremely important concept, called the PSD which will be discussed in the next handout.
- It will be proved that the PSD should always be positive.
- It is easy to prove that an ACS or ACF has a positive Fourier transform if, and only if, it is positive semi-definite.

Example. Consider the following functions. For each function, state whether it is a valid autocorrelation function or autocorrelation sequence or not. Explain carefully the reasoning for your answers, but no detailed calculations are required.

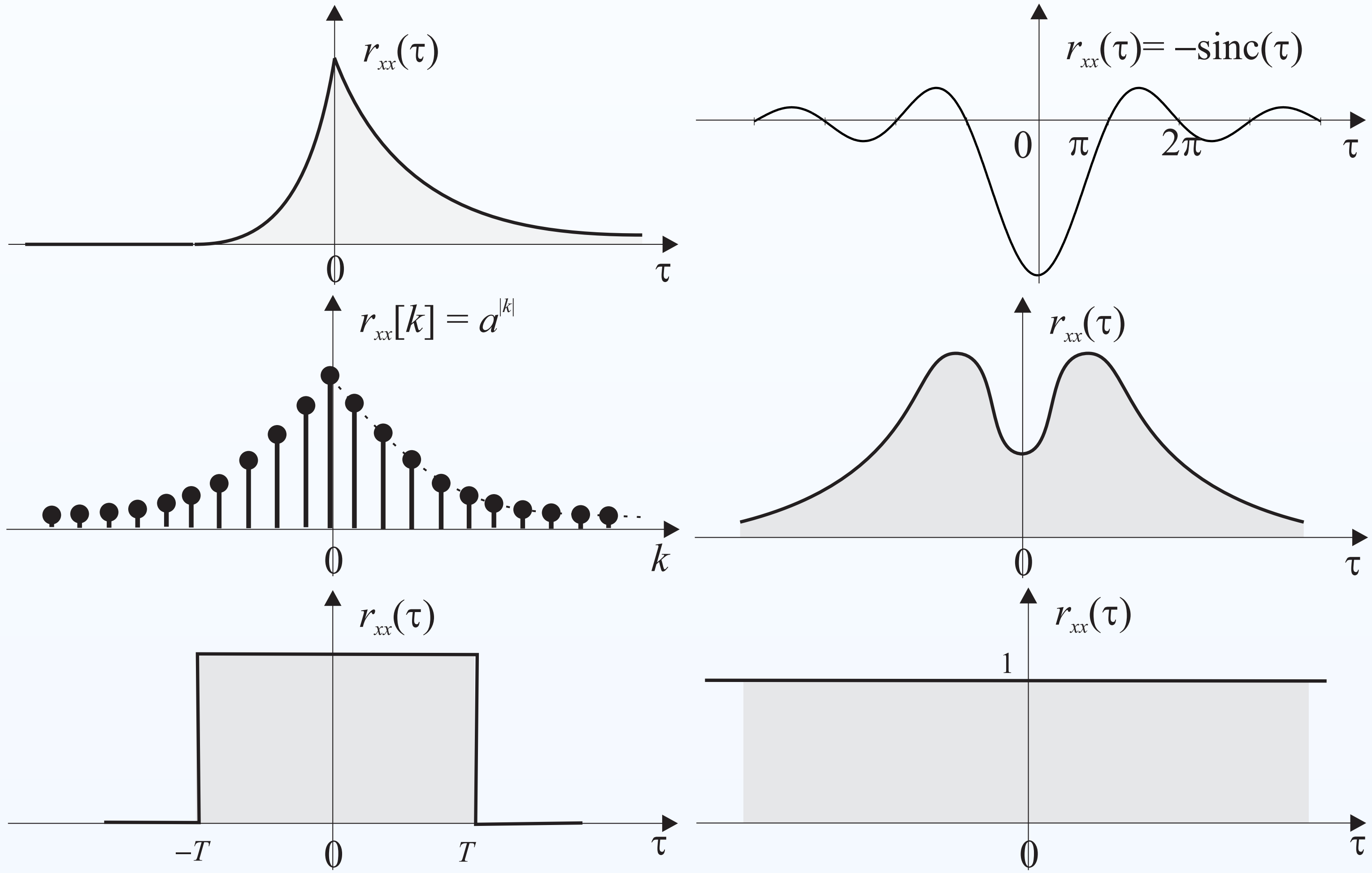
WSS Properties



Candidate autocorrelation functions.

Example. SOLUTION. For each function, test the four properties.

WSS Properties



Candidate autocorrelation functions.

Example. SOLUTION. Thus: 1)-2) and 4)-5), No; 3) and 6) Yes!

WSS Properties

– End-of-Topic 49: **Properties of the ACS for WSS** –

Any Questions?

Wide-sense cyclo-stationarity

- A signal whose statistical properties vary *cyclically* with time is called a cyclostationary process.

Wide-sense cyclo-stationarity

- A signal whose statistical properties vary *cyclically* with time is called a cyclostationary process.
- A cyclostationary process can be viewed as several interleaved stationary processes.

Wide-sense cyclo-stationarity

- A signal whose statistical properties vary *cyclically* with time is called a cyclostationary process.
- A cyclostationary process can be viewed as several interleaved stationary processes.
- For example, the maximum daily temperature in Edinburgh can be modeled as a cyclostationary process: the maximum temperature on July 21 is statistically different from the temperature on December 18; however, the temperature on December 18 of different years has (arguably) identical statistics.

Wide-sense cyclo-stationarity

- A signal whose statistical properties vary *cyclically* with time is called a cyclostationary process.
- A cyclostationary process can be viewed as several interleaved stationary processes.
- For example, the maximum daily temperature in Edinburgh can be modeled as a cyclostationary process: the maximum temperature on July 21 is statistically different from the temperature on December 18; however, the temperature on December 18 of different years has (arguably) identical statistics.
- Two classes of **nonstationary process** which, in part, have properties resembling stationary signals are:

Wide-sense cyclo-stationarity

- Two classes of **nonstationary process** which, in part, have properties resembling stationary signals are:

1. **A wide-sense periodic (WSP) process** is classified as signals whose mean is periodic, and whose ACS is periodic in both dimensions:

$$\mu_x[n] = \mu_x[n+N]$$

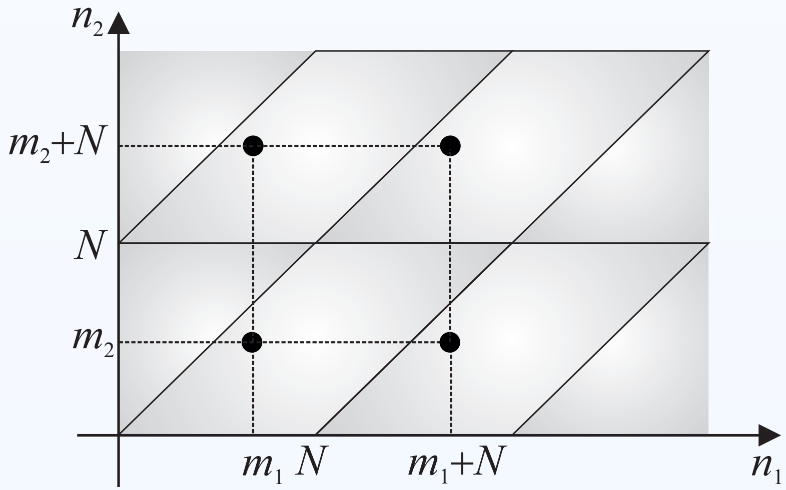
$$\begin{aligned} r_{xx}[n_1, n_2] &= r_{xx}[n_1+N, n_2] = r_{xx}[n_1, n_2+N] \\ &= r_{xx}[n_1+N, n_2+N] \end{aligned}$$

for all n, n_1 and n_2. These are quite tight constraints.

Wide-sense cyclo-stationarity

- Two classes of **nonstationary process** which, in part, have properties resembling stationary signals are:

1. **A WSP process** is classified as signals whose mean is periodic, and whose ACS is periodic in both dimensions:



The periodicity of the ACS for a WSP signal.

Wide-sense cyclo-stationarity

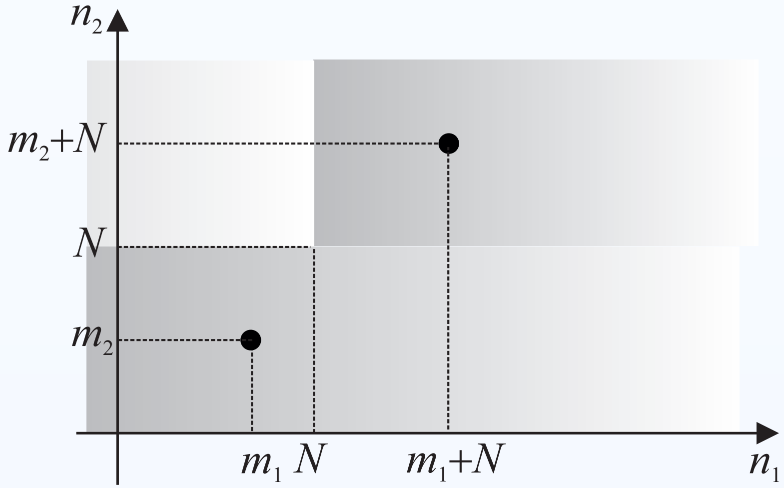
2. **A wide-sense cyclo-stationary process** has similar but less restrictive properties than a WSP process, in that the mean is periodic, but the ACS is now just invariant to a shift by N in both of its arguments:

$$\mu_x[n] = \mu_x[n+N]$$
$$r_{xx}[n_1, n_2] = r_{xx}[n_1 + N, n_2 + N]$$

for all n, n_1 and n_2. This type of nonstationary process has more practical applications.

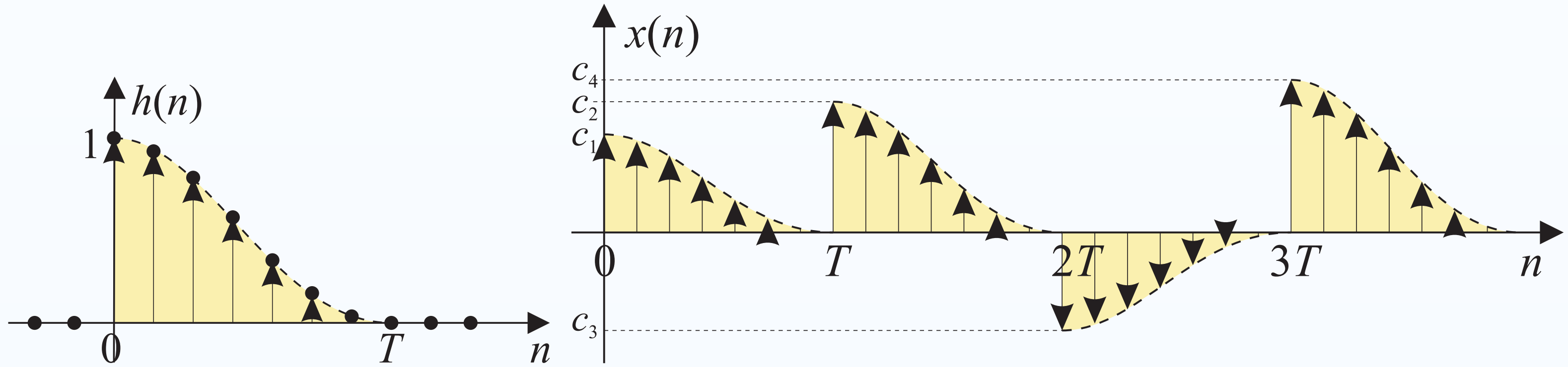
Wide-sense cyclo-stationarity

2. **A wide-sense cyclo-stationary process** has similar but less restrictive properties than a WSP process, in that the mean is periodic, but the ACS is now just invariant to a shift by N in both of its arguments:



The periodicity of the ACS for a wide-sense cyclo-stationary process.

Wide-sense cyclo-stationarity



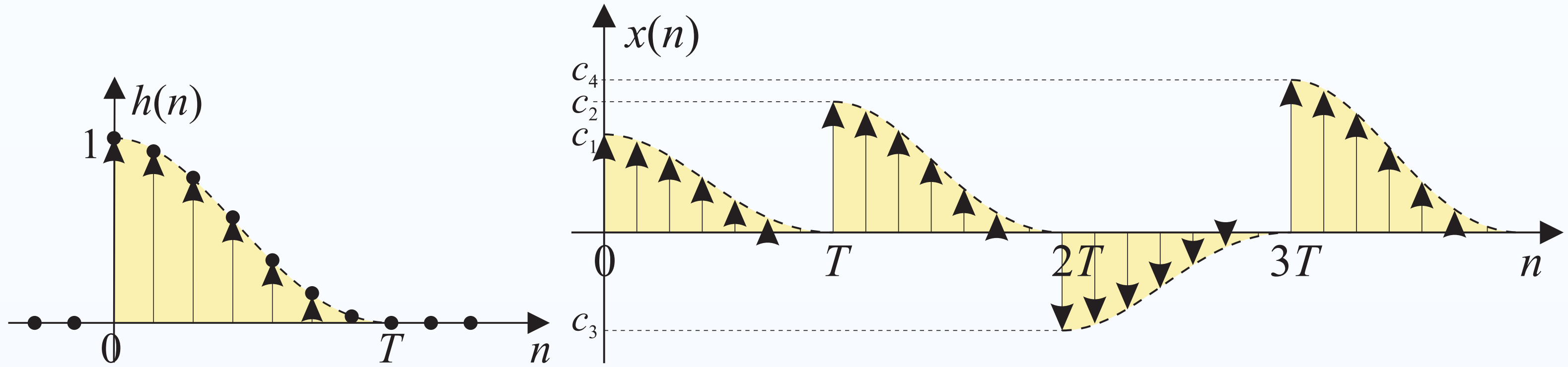
An example pulse and typical transmit signal.

Example (Pulse-Amplitude Modulation). An important example of a cyclo-stationary process is the random signal:

$$x[n] = \sum_{m=-\infty}^{\infty} c_m \, h[n - mT] \qquad \bowtie$$

for some period T, where c_m is a stationary sequence with ACS $r_{cc}[n_1, n_2] = \mathbb{E}\left[c_{n_1} \, c^*_{n_2}\right] = r_{cc}[n_1 - n_2]$, and $h[n]$ is a given deterministic sequence, usually an impulse response.

Wide-sense cyclo-stationarity



An example pulse and typical transmit signal.

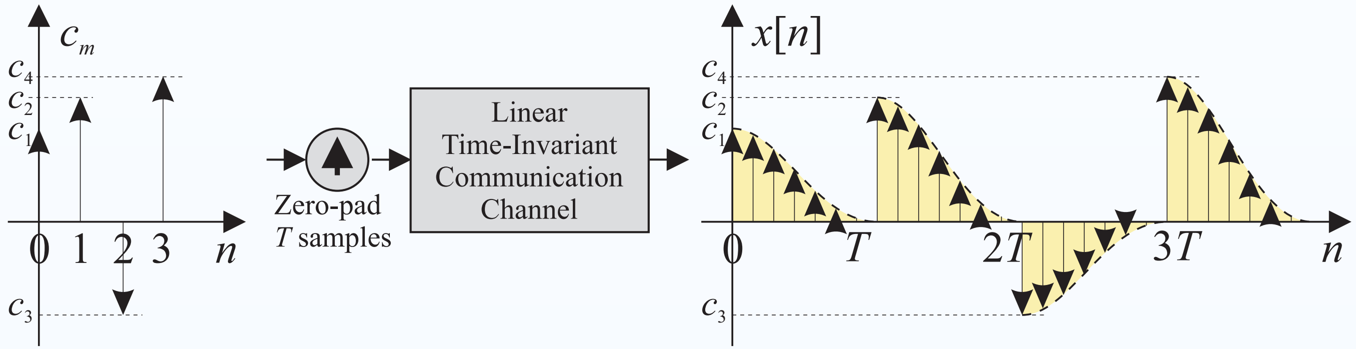
Example (Pulse-Amplitude Modulation). An important example of a cyclo-stationary process is the random signal:

$$x[n] = \sum_{m=-\infty}^{\infty} c_m \, h[n - mT]$$

⋈

Show that $x[n]$ satisfies the properties of a wide-sense cyclo-stationary process.

Wide-sense cyclo-stationarity



Example (Pulse-Amplitude Modulation). An important example of a cyclo-stationary process is the random signal:

$$x[n] = \sum_{m=-\infty}^{\infty} c_m \, h[n - mT]$$

SOLUTION.
- $x[n]$ represents the signal for several different types of linear modulation used in digital communications.
- $\{c_m\}$ represents the digital information that is transmitted over the communication channel, and $\frac{1}{T}$ represents the rate of transmission of the information symbols. □

Wide-sense cyclo-stationarity

Example (Pulse-Amplitude Modulation). An important example of a cyclo-stationary process is the random signal:

$$x[n] = \sum_{m=-\infty}^{\infty} c_m \, h[n - mT]$$

SOLUTION. To see this is wide-sense cyclo-stationary:

$$\mu_x[n] = \mathbb{E}\left[x[n]\right] = \sum_{m=-\infty}^{\infty} \mathbb{E}\left[c_m\right] h[n - mT] = \mu_c \sum_{m=-\infty}^{\infty} h[n - mT]$$

□

where $\mu_c[n] = \mu_c$ since it is a stationary process.

Wide-sense cyclo-stationarity

Example (Pulse-Amplitude Modulation). An important example of a cyclo-stationary process is the random signal:

$$x[n] = \sum_{m=-\infty}^{\infty} c_m \, h[n - mT]$$

SOLUTION. To see this is wide-sense cyclo-stationary:

$$\mu_x[n] = \mathbb{E}\left[x[n]\right] = \sum_{m=-\infty}^{\infty} \mathbb{E}\left[c_m\right] h[n - mT] = \mu_c \sum_{m=-\infty}^{\infty} h[n - mT]$$

where $\mu_c[n] = \mu_c$ since it is a stationary process. Thus, observe:

$$\mu_x[n + kT] = \mu_c \sum_{m=-\infty}^{\infty} h[n + kT - Tm] = \mu_c \sum_{r=-\infty}^{\infty} h[n - Tr] = \mu_x[n]$$

□

Wide-sense cyclo-stationarity

Example (Pulse-Amplitude Modulation). An important example of a cyclo-stationary process is the random signal:

$$x[n] = \sum_{m=-\infty}^{\infty} c_m \, h[n - mT]$$

SOLUTION. Next consider the autocorrelation function given by:

$$\begin{aligned} r_{xx}[n_1, n_2] &= \mathbb{E}\left[x[n_1] \; x^*[n_2]\right] \\ &= \sum_{m=-\infty}^{\infty} \sum_{\ell=-\infty}^{\infty} h[n_1 - Tm] \; h[n_2 - T\ell] \; r_{cc}[m - \ell] \end{aligned}$$ □

where $r_{cc}[m, \ell] = \mathbb{E}\left[c_m \, c_\ell^*\right] = r_{cc}[m - \ell]$ since it is a stationary process.

Wide-sense cyclo-stationarity

Example (Pulse-Amplitude Modulation). An important example of a cyclo-stationary process is the random signal:

$$x[n] = \sum_{m=-\infty}^{\infty} c_m \, h[n - mT]$$

SOLUTION. Next consider the autocorrelation function given by:

$$r_{xx}[n_1, n_2] = \mathbb{E}\left[x[n_1] \; x^*[n_2]\right]$$

$$= \sum_{m=-\infty}^{\infty} \sum_{\ell=-\infty}^{\infty} h[n_1 - Tm] \; h[n_2 - T\ell] \; r_{cc}[m - \ell] \quad \square$$

where $r_{cc}[m, \ell] = \mathbb{E}\left[c_m \, c_\ell^*\right] = r_{cc}[m - \ell]$ since it is a stationary process. Similar to the approach above, then set $n_1 \to n_1 + pT$ and $n_2 \to n_2 + qT$.

Wide-sense cyclo-stationarity

Example (Pulse-Amplitude Modulation). An important example of a cyclo-stationary process is the random signal:

$$x[n] = \sum_{m=-\infty}^{\infty} c_m \, h[n - mT]$$

SOLUTION. Therefore, it follows:

$$\begin{aligned} &r_{xx}[n_1 + pT, \, n_2 + qT] \\ &= \sum_{m=-\infty}^{\infty} \sum_{\ell=-\infty}^{\infty} h[n_1 - T(m-p)] \; h[n_2 - T(l-q)] \; r_{cc}[m-\ell] \end{aligned}$$

□

Wide-sense cyclo-stationarity

Example (Pulse-Amplitude Modulation). An important example of a cyclo-stationary process is the random signal:

$$x[n] = \sum_{m=-\infty}^{\infty} c_m \, h[n - mT]$$

SOLUTION. Again, by setting $r = m - p$ and $s = \ell - q$:

$$r_{xx}[n_1 + pT, n_2 + qT]$$
$$= \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} h[n_1 - Tr] \; h[n_2 - Ts] \; r_{cc}[r - s + p - q]$$

□

Wide-sense cyclo-stationarity

Example (Pulse-Amplitude Modulation). An important example of a cyclo-stationary process is the random signal:

$$x[n] = \sum_{m=-\infty}^{\infty} c_m \, h[n - mT]$$

SOLUTION. Again, by setting $r = m - p$ and $s = \ell - q$:

$$r_{xx}[n_1 + pT, n_2 + qT]$$
$$= \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} h[n_1 - Tr] \; h[n_2 - Ts] \; r_{cc}[r - s + p - q]$$

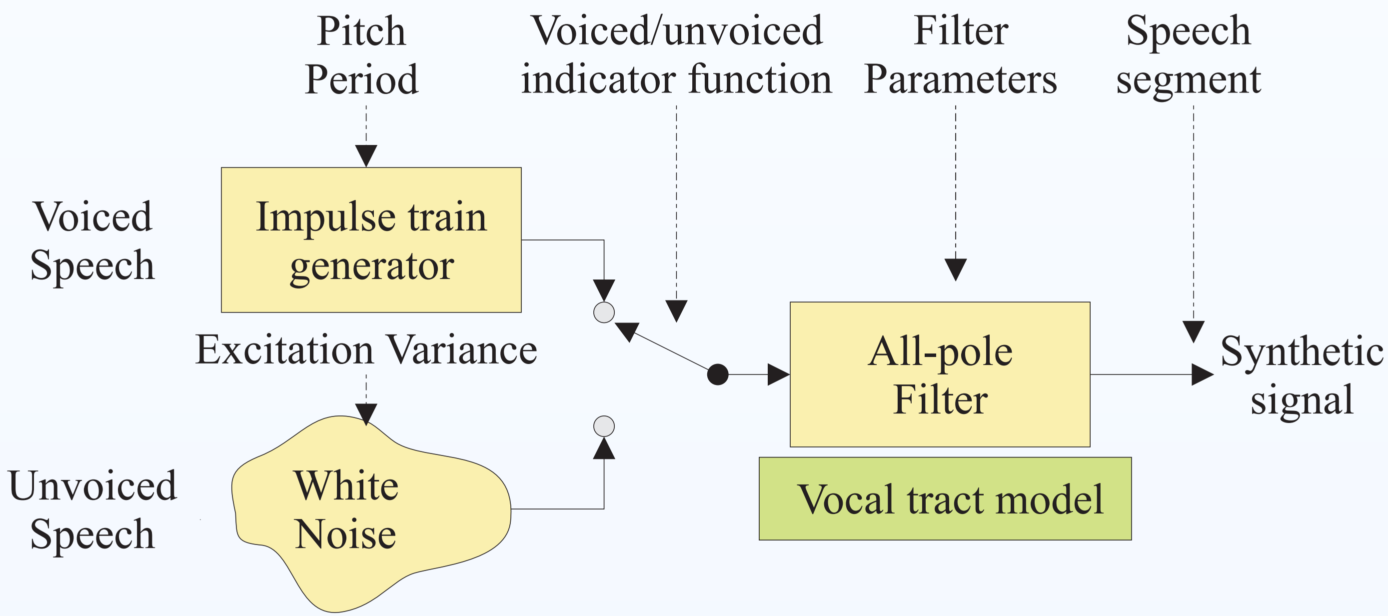
In the case that $p = q$, then it finally follows that:

$$r_{xx}[n_1 + pT, n_2 + pT] = r_{xx}[n_1, n_2] \qquad \square$$

By definition, $x[n]$ is therefore a cyclo-stationary process.

Quasi-stationarity

At the introduction of this lecture course, it was noted that in the analysis of speech signals, the speech waveform is broken up into short segments whose duration is typically 10 to 20 milliseconds.



The speech synthesis model (repeated from Introduction handout).

Quasi-stationarity

At the introduction of this lecture course, it was noted that in the analysis of speech signals, the speech waveform is broken up into short segments whose duration is typically 10 to 20 milliseconds.

This is because speech can be modelled as a **locally stationary** or **quasi-stationary** process.

Quasi-stationarity

At the introduction of this lecture course, it was noted that in the analysis of speech signals, the speech waveform is broken up into short segments whose duration is typically 10 to 20 milliseconds.

This is because speech can be modelled as a **locally stationary** or **quasi-stationary** process.

- Such processes possess statistical properties that change *slowly* over short periods of time.

Quasi-stationarity

At the introduction of this lecture course, it was noted that in the analysis of speech signals, the speech waveform is broken up into short segments whose duration is typically 10 to 20 milliseconds.

This is because speech can be modelled as a **locally stationary** or **quasi-stationary** process.

- Such processes possess statistical properties that change *slowly* over short periods of time.
- They are *globally* nonstationary, but are approximately *locally* stationary, and are modelled as if the statistics *actually are* stationary over a short segment of time.

Quasi-stationarity

At the introduction of this lecture course, it was noted that in the analysis of speech signals, the speech waveform is broken up into short segments whose duration is typically 10 to 20 milliseconds.

This is because speech can be modelled as a **locally stationary** or **quasi-stationary** process.

- Such processes possess statistical properties that change *slowly* over short periods of time.
- They are *globally* nonstationary, but are approximately *locally* stationary, and are modelled as if the statistics *actually are* stationary over a short segment of time.
- Quasi-stationary models are, in fact, just a special case of nonstationary processes, but are distinguished since their characterisation closely resemble stationary processes.

Quasi-stationarity

– End-of-Topic 50: **Wide-sense periodic and cyclostationary signals, and other forms of nonstationary signals** –

Any Questions?

Estimating statistical properties

- A stochastic process consists of the ensemble, $x[n, \zeta]$, and a probability law, $f_X(\{x\} \mid \{n\})$. If this information is available $\forall n$, the statistical properties are easily determined.

Estimating statistical properties

- A stochastic process consists of the ensemble, $x[n, \zeta]$, and a probability law, $f_X\left(\{x\} \mid \{n\}\right)$. If this information is available $\forall n$, the statistical properties are easily determined.

- In practice, only a limited number of realisations of a process is available, and often only one: i.e. $\{x[n, \zeta_k]\,,\, k \in \{1, \dots, K\}\}$ is known for some K, but $f_X\left(x \mid n\right)$ is unknown.

Estimating statistical properties

- A stochastic process consists of the ensemble, $x[n, \zeta]$, and a probability law, $f_X(\{x\} \mid \{n\})$. If this information is available $\forall n$, the statistical properties are easily determined.

- In practice, only a limited number of realisations of a process is available, and often only one: i.e. $\{x[n, \zeta_k], k \in \{1, \dots, K\}\}$ is known for some K, but $f_X(x \mid n)$ is unknown.

- Is is possible to infer the statistical characteristics of a process from a single realisation? Yes, for the following class of signals:

Estimating statistical properties

- A stochastic process consists of the ensemble, $x[n, \zeta]$, and a probability law, $f_X(\{x\} \mid \{n\})$. If this information is available $\forall n$, the statistical properties are easily determined.

- In practice, only a limited number of realisations of a process is available, and often only one: i.e. $\{x[n, \zeta_k], k \in \{1, \dots, K\}\}$ is known for some K, but $f_X(x \mid n)$ is unknown.

- Is is possible to infer the statistical characteristics of a process from a single realisation? Yes, for the following class of signals:

  - **ergodic** processes;

Estimating statistical properties

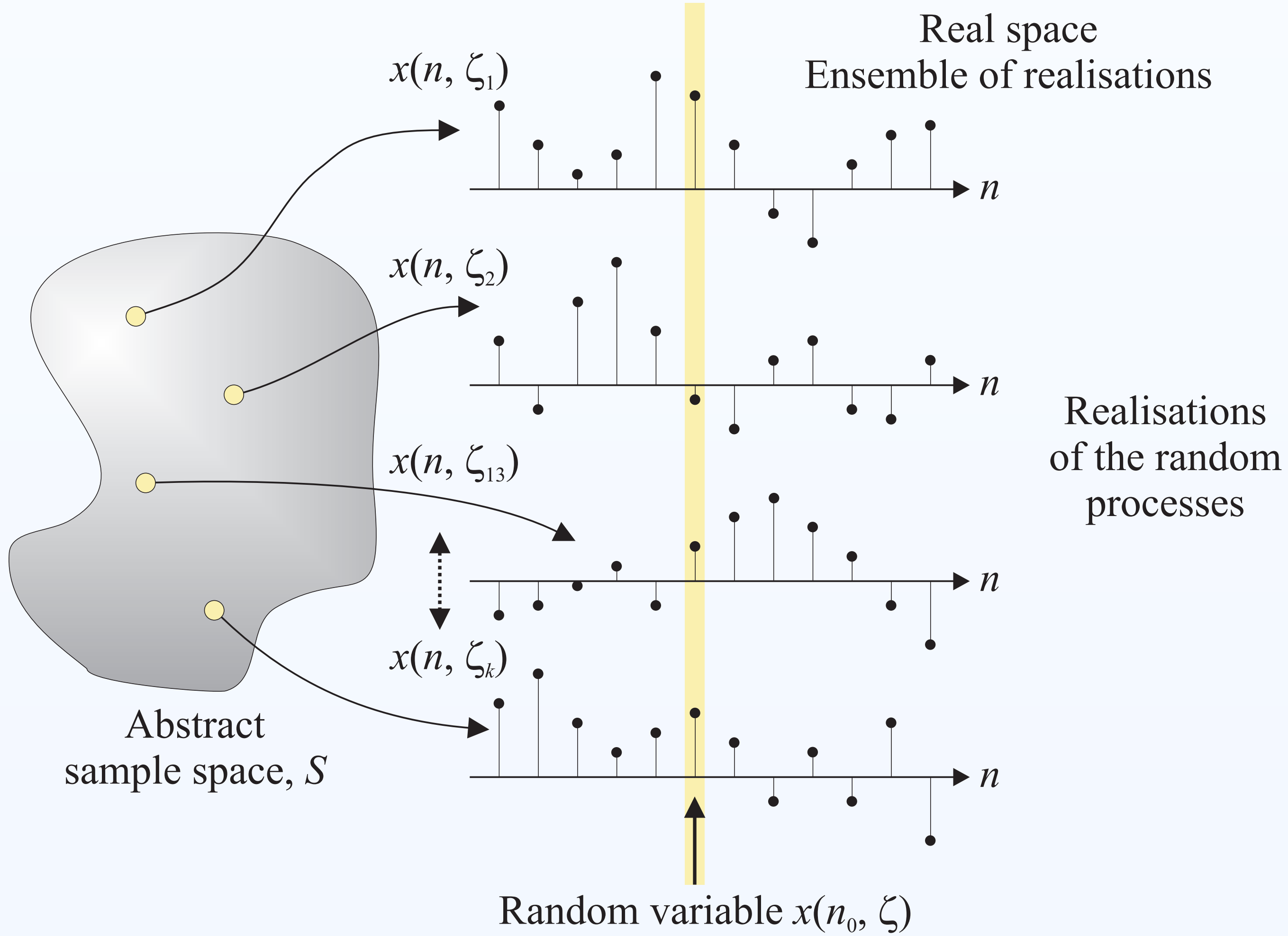
- A stochastic process consists of the ensemble, $x[n, \zeta]$, and a probability law, $f_X(\{x\} \mid \{n\})$. If this information is available $\forall n$, the statistical properties are easily determined.

- In practice, only a limited number of realisations of a process is available, and often only one: i.e. $\{x[n, \zeta_k], k \in \{1, \dots, K\}\}$ is known for some K, but $f_X(x \mid n)$ is unknown.

- Is is possible to infer the statistical characteristics of a process from a single realisation? Yes, for the following class of signals:
  - **ergodic** processes;
  - nonstationary processes where additional structure about the autocorrelation function is known (beyond the scope of this course).

Ensemble and Time-Averages

Ensemble averaging, as considered so far in the course, is not frequently used in practice since it is impractical to obtain the number of realisations needed for an accurate estimate.



Ensemble and Time-Averages

Ensemble averaging, as considered so far in the course, is not frequently used in practice since it is impractical to obtain the number of realisations needed for an accurate estimate.

A statistical average that can be obtained from a **single** realisation of a process is a **time-average**, defined by:

$$\langle g(x[n]) \rangle \triangleq \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} g(x[n])$$

- For every ensemble average, a corresponding time-average can be defined; the above corresponds to: $\mathbb{E}[g(x[n])]$.

Ensemble and Time-Averages

Ensemble averaging, as considered so far in the course, is not frequently used in practice since it is impractical to obtain the number of realisations needed for an accurate estimate.

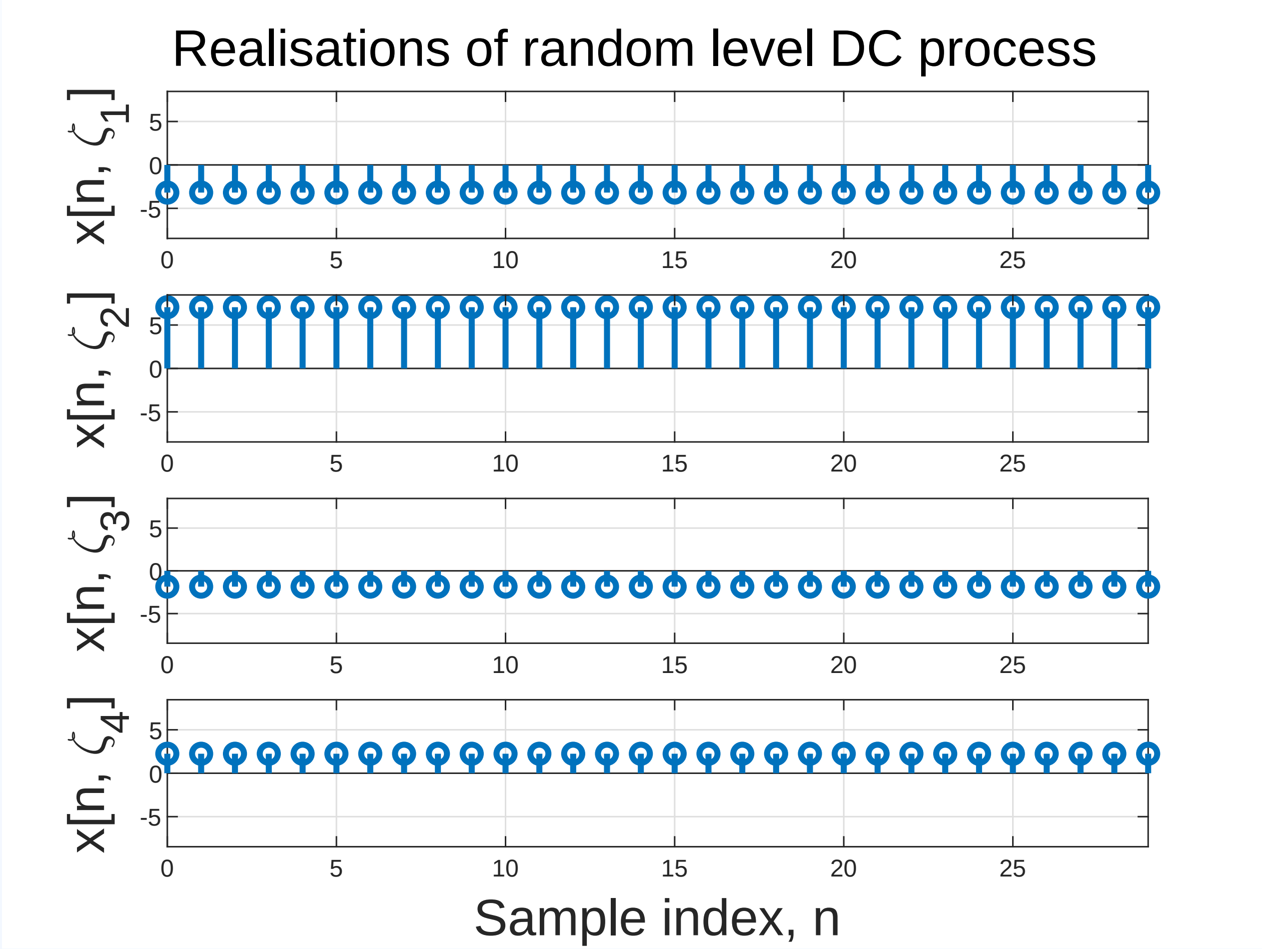
A statistical average that can be obtained from a **single** realisation of a process is a **time-average**, defined by:

$$\langle g(x[n]) \rangle \triangleq \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} g(x[n])$$

- For every ensemble average, a corresponding time-average can be defined; the above corresponds to: $\mathbb{E}\left[g(x[n])\right]$.
- Time-averages are random variables since they implicitly depend on the particular realisation, given by ζ. Averages of deterministic signals are fixed numbers or sequences, even though they are given by the same expression.

Ensemble and Time-Averages

- Ergodicity requires a single realisation of the random process to display the behaviour of the entire ensemble of realisations.



Ergodicity

A stochastic process, $x[n]$, is **ergodic** if its ensemble averages can be estimated from a single realisation of a process using time averages.

The two most important degrees of ergodicity are:

Mean-Ergodic (or ergodic in the mean) processes have identical expected values and sample-means:

$$\langle x[n] \rangle = \mathbb{E}\left[x[n]\right]$$

Covariance-Ergodic Processes (or ergodic in correlation) have the property that:

$$\langle x[n]\, x^*[n-l] \rangle = \mathbb{E}\left[x[n]\, x^*[n-l]\right]$$

Ergodicity

- It should be intuitiveness obvious that ergodic processes must be stationary and, moreover, that a process which is ergodic both in the mean and correlation is WSS.
- WSS processes are not necessarily ergodic.
- Ergodic is often used to mean both ergodic in the mean and correlation.
- In practice, only finite records of data are available, and therefore an estimate of the time-average will be given by

$$\langle g(x[n]) \rangle = \frac{1}{N} \sum_{n \in \mathcal{N}} g(x[n])$$

where N is the number of data-points available. The performance of this estimator will be discussed elsewhere in this course.

More Details on Mean-Ergodicity

The **time-average** over $2N+1$ samples, $\{x[n]\}_{-N}^{N}$ is:

$$\mu_x|_N = \langle x[n] \rangle = \frac{1}{2N+1} \sum_{n=-N}^{N} x[n]$$

More Details on Mean-Ergodicity

The **time-average** over $2N+1$ samples, $\{x[n]\}_{-N}^{N}$ is:

$$\mu_x|_N = \langle x[n] \rangle = \frac{1}{2N+1} \sum_{n=-N}^{N} x[n]$$

$\mu_X|_N$ is a random variable with mean:

$$\mathbb{E}\left[\mu_x|_N\right] = \frac{1}{2N+1} \sum_{n=-N}^{N} \mathbb{E}\left[x[n]\right] = \mu_x$$

This is an **unbiased estimate**.

More Details on Mean-Ergodicity

The **time-average** over $2N+1$ samples, $\{x[n]\}_{-N}^{N}$ is:

$$\mu_x|_N = \langle x[n] \rangle = \frac{1}{2N+1} \sum_{n=-N}^{N} x[n]$$

$\mu_X|_N$ is a random variable with mean:

$$\mathbb{E}\left[\mu_x|_N\right] = \frac{1}{2N+1} \sum_{n=-N}^{N} \mathbb{E}\left[x[n]\right] = \mu_x$$

This is an **unbiased estimate.**

Since $\mu_x|_N$ is a random variable, then it must have a variance:

$$\operatorname{var}\left[\mu_x|_N\right] = \operatorname{var}\left[\frac{1}{2N+1} \sum_{n=-N}^{N} x[n]\right]$$

More Details on Mean-Ergodicity

The **time-average** over $2N+1$ samples, $\{x[n]\}_{-N}^{N}$ is:

$$\mu_x|_N = \langle x[n] \rangle = \frac{1}{2N+1} \sum_{n=-N}^{N} x[n]$$

Theorem (Variance of estimator). If $x[n]$ has ACS $\gamma_{xx}[\ell]$, then:

$$\mathrm{var}\left[\mu_x|_N\right] = \frac{1}{2N+1} \sum_{\ell=-2N}^{2N} \left(1 - \frac{|\ell|}{2N+1}\right) \gamma_{xx}[\ell] \qquad \diamondsuit$$

More Details on Mean-Ergodicity

The **time-average** over $2N+1$ samples, $\{x[n]\}_{-N}^{N}$ is:

$$\mu_x|_N = \langle x[n] \rangle = \frac{1}{2N+1} \sum_{n=-N}^{N} x[n]$$

Theorem (Variance of estimator). If $x[n]$ has ACS $\gamma_{xx}[\ell]$, then:

$$\text{var}\left[\mu_x|_N\right] = \frac{1}{2N+1} \sum_{\ell=-2N}^{2N} \left(1 - \frac{|\ell|}{2N+1}\right) \gamma_{xx}[\ell] \qquad \diamondsuit$$

- If $\lim_{N\to\infty} \text{var}\left[\mu_x|_N\right] = 0$, then $\mu_x|_N \to \mu_x$ in the mean-square sense.
- It is said that the time average $\mu_x|_N$ computed from a single realisation of $x[n]$ is close to μ_x with probability close to 1.
- If this is true, then the process $x[n]$ is **mean-ergodic**.

More Details on Mean-Ergodicity

The **time-average** over $2N+1$ samples, $\{x[n]\}_{-N}^{N}$ is:

$$\mu_x|_N = \langle x[n] \rangle = \frac{1}{2N+1} \sum_{n=-N}^{N} x[n]$$

- The result presented above leads to the following conclusion:

Theorem (Mean-ergodic processes). A discrete-random process $x[n]$ with autocovariance $\gamma_{xx}[\ell]$ is mean-ergodic iff:

$$\lim_{N\to\infty} \frac{1}{2N+1} \sum_{\ell=-2N}^{2N} \left(1 - \frac{|\ell|}{2N+1}\right) \gamma_{xx}[\ell] = 0$$

PROOF. See discussion above.

More Details on Mean-Ergodicity

Example ([Papoulis:1991, Example 13.3, Page 429]). A stationary stochastic process $x[n]$ has an ACS given by $\gamma_{xx}[\ell] = q\, e^{-c\,|\ell|}$ for some constants q and c. Is $x[n]$ ergodic in the mean?

More Details on Mean-Ergodicity

Example ([Papoulis:1991, Example 13.3, Page 429]). A stationary stochastic process $x[n]$ has an ACS given by $\gamma_{xx}[\ell] = q\, e^{-c\,|\ell|}$ for some constants q and c. Is $x[n]$ ergodic in the mean?

SOLUTION. Writing:

$$\operatorname{var}\left[\mu_x|_N\right] = \frac{1}{2N+1} \sum_{\ell=-2N}^{2N} \left(1 - \frac{|\ell|}{2N+1}\right) \gamma_{xx}[\ell]$$

$$= \frac{q}{2N+1} \sum_{\ell=-2N}^{2N} \left(1 - \frac{|\ell|}{2N+1}\right) e^{-c|\ell|}$$

□

More Details on Mean-Ergodicity

Example ([Papoulis:1991, Example 13.3, Page 429]). A stationary stochastic process $x[n]$ has an ACS given by $\gamma_{xx}[\ell] = q\, e^{-c\,|\ell|}$ for some constants q and c. Is $x[n]$ ergodic in the mean?

SOLUTION. Writing:

$$\operatorname{var}\left[\mu_x|_N\right] = \frac{1}{2N+1}\sum_{\ell=-2N}^{2N}\left(1-\frac{|\ell|}{2N+1}\right)\gamma_{xx}[\ell]$$

$$= \frac{q}{2N+1}\sum_{\ell=-2N}^{2N}\left(1-\frac{|\ell|}{2N+1}\right)e^{-c|\ell|}$$

which can be rearranged to give as:

$$\operatorname{var}\left[\mu_x|_N\right] = \frac{q}{2N+1}\left\{2\sum_{\ell=0}^{2N}\left(1-\frac{\ell}{2N+1}\right)e^{-c\,\ell}-1\right\}$$

More Details on Mean-Ergodicity

Example ([Papoulis:1991, Example 13.3, Page 429]). A stationary stochastic process $x[n]$ has an ACS given by $\gamma_{xx}[\ell] = q\, e^{-c\,|\ell|}$ for some constants q and c. Is $x[n]$ ergodic in the mean?

SOLUTION. Writing:

$$\mathrm{var}\left[\mu_x|_N\right] = \frac{q}{2N+1}\left\{2\sum_{\ell=0}^{2N}\left(1-\frac{\ell}{2N+1}\right)e^{-c\,\ell} - 1\right\}$$

Now, noting the general result:

$$\sum_{n=0}^{N-1}(a+nb)r^n = \frac{a-[a+(N-1)b]r^N}{1-r} + \frac{br(1-r^{N-1})}{(1-r)^2}$$

□

More Details on Mean-Ergodicity

Example ([Papoulis:1991, Example 13.3, Page 429]). A stationary stochastic process $x[n]$ has an ACS given by $\gamma_{xx}[\ell] = q\, e^{-c\,|\ell|}$ for some constants q and c. Is $x[n]$ ergodic in the mean?

Solution. Writing:

$$\text{var}\left[\mu_x|_N\right] = \frac{q}{2N+1}\left\{2\sum_{\ell=0}^{2N}\left(1-\frac{\ell}{2N+1}\right)e^{-c\,\ell} - 1\right\}$$

Now, noting the general result:

$$\sum_{n=0}^{N-1}(a+nb)r^n = \frac{a-[a+(N-1)b]r^N}{1-r} + \frac{br(1-r^{N-1})}{(1-r)^2} \qquad \square$$

then by setting $a=1$, $b=-\frac{1}{2N+1}$ and $r=e^{-c}$, with $n=\ell$ and $N \rightarrow 2N+1$:

More Details on Mean-Ergodicity

Example ([Papoulis:1991, Example 13.3, Page 429]). A stationary stochastic process $x[n]$ has an ACS given by $\gamma_{xx}[\ell] = q\, e^{-c\,|\ell|}$ for some constants q and c. Is $x[n]$ ergodic in the mean?

SOLUTION. Writing:

$$\operatorname{var}\left[\mu_x|_N\right] = \frac{q}{2N+1}\left\{2\sum_{\ell=0}^{2N}\left(1-\frac{\ell}{2N+1}\right)e^{-c\,\ell} - 1\right\}$$

the variance can be written as (where $M = 2N+1$):

$$\operatorname{var}\left[\mu_x|_N\right] = 2q\left[\frac{\frac{1}{M} - \frac{1}{M^2}e^{-Mc}}{1-e^{-c}} + \frac{\frac{1}{M^2}e^{-c} - \frac{1}{M^2}e^{-Mc}}{(1-e^{-c})^2} - \frac{1}{2M}\right]$$

□

More Details on Mean-Ergodicity

Example ([Papoulis:1991, Example 13.3, Page 429]). A stationary stochastic process $x[n]$ has an ACS given by $\gamma_{xx}[\ell] = q\, e^{-c\,|\ell|}$ for some constants q and c. Is $x[n]$ ergodic in the mean?

SOLUTION. Writing:

$$\operatorname{var}\left[\mu_x|_N\right] = \frac{q}{2N+1}\left\{2\sum_{\ell=0}^{2N}\left(1-\frac{\ell}{2N+1}\right)e^{-c\,\ell} - 1\right\}$$

the variance can be written as (where $M = 2N+1$):

$$\operatorname{var}\left[\mu_x|_N\right] = 2q\left[\frac{\frac{1}{M} - \frac{1}{M^2}e^{-Mc}}{1-e^{-c}} + \frac{\frac{1}{M^2}e^{-c} - \frac{1}{M^2}e^{-Mc}}{(1-e^{-c})^2} - \frac{1}{2M}\right]$$

Now, by setting $N \to \infty$, which is equivalent to $M \to \infty$, and:

$$\lim_{n\to\infty} n^s\, x^n \to 0 \quad \text{if } |x| < 1 \text{ for any real value of } s$$

More Details on Mean-Ergodicity

Example ([Papoulis:1991, Example 13.3, Page 429]). A stationary stochastic process $x[n]$ has an ACS given by $\gamma_{xx}[\ell] = q\,e^{-c\,|\ell|}$ for some constants q and c. Is $x[n]$ ergodic in the mean?

SOLUTION. Writing:

$$\text{var}\left[\mu_x|_N\right] = \frac{q}{2N+1}\left\{2\sum_{\ell=0}^{2N}\left(1-\frac{\ell}{2N+1}\right)e^{-c\,\ell} - 1\right\}$$

the variance can be written as (where $M = 2N+1$):

$$\text{var}\left[\mu_x|_N\right] = 2q\left[\frac{\frac{1}{M}-\frac{1}{M^2}e^{-Mc}}{1-e^{-c}} + \frac{\frac{1}{M^2}e^{-c}-\frac{1}{M^2}e^{-Mc}}{(1-e^{-c})^2} - \frac{1}{2M}\right]$$

it can be seen that since $M = 2N+1$:

$$\lim_{N\to\infty}\text{var}\left[\mu_x|_N\right] = 0 \qquad \square$$

More Details on Mean-Ergodicity

– End-of-Topic 51: **Ergodicity and time-average estimates of statistics of WSS processes** –

Any Questions?

Joint Signal Statistics

Cross-correlation and cross-covariance A measure of the dependence between values of two *different* stochastic processes is given by the **cross-correlation** and **cross-covariance** functions:

$$r_{xy}[n_1, n_2] = \mathbb{E}\left[x[n_1]\; y^*[n_2]\right]$$

$$\gamma_{xy}[n_1, n_2] = r_{xy}[n_1, n_2] - \mu_x[n_1]\; \mu_y^*[n_2]$$

Joint Signal Statistics

Cross-correlation and cross-covariance A measure of the dependence between values of two *different* stochastic processes is given by the **cross-correlation** and **cross-covariance** functions:

$$r_{xy}[n_1, n_2] = \mathbb{E}\left[x[n_1]\; y^*[n_2]\right]$$
$$\gamma_{xy}[n_1, n_2] = r_{xy}[n_1, n_2] - \mu_x[n_1]\; \mu_y^*[n_2]$$

Normalised cross-correlation (or cross-covariance) The cross-covariance provides a measure of similarity of the deviation from the respective means of two processes. It makes sense to consider this deviation relative to their **standard deviations**; thus, **normalised cross-correlations**:

$$\rho_{xy}[n_1, n_2] = \frac{\gamma_{xy}[n_1, n_2]}{\sigma_x[n_1]\; \sigma_y[n_2]}$$

Types of Joint Stochastic Processes

Statistically independence of two stochastic processes occurs when, for every n_x and n_y,

$$f_{XY}(x, y \mid n_x, n_y) = f_X(x \mid n_x)\; f_Y(y \mid n_y)$$

Uncorrelated stochastic processes have, for all n_x & $n_y \neq n_x$:

$$\gamma_{xy}[n_x, n_y] = 0$$
$$r_{xy}[n_x, n_y] = \mu_x[n_x]\; \mu_y[n_y]$$

Types of Joint Stochastic Processes

Statistically independence of two stochastic processes occurs when, for every n_x and n_y,

$$f_{XY}\left(x, y \,|\, n_x, n_y\right) = f_X\left(x \,|\, n_x\right)\, f_Y\left(y \,|\, n_y\right)$$

Uncorrelated stochastic processes have, for all n_x & $n_y \neq n_x$:

$$\gamma_{xy}[n_x, n_y] = 0$$

$$r_{xy}[n_x, n_y] = \mu_x[n_x]\, \mu_y[n_y]$$

Joint stochastic processes that are statistically independent are uncorrelated, but not necessarily vice-versa, except for Gaussian processes.

Types of Joint Stochastic Processes

Orthogonal joint processes have, for every n_1 and $n_2 \neq n_1$:

$$r_{xy}[n_1, n_2] = 0$$

Joint WSS is a similar to WSS for a single stochastic process, and is useful since it facilitates a spectral description, as discussed later in this course:

$$r_{xy}[\ell] = r_{xy}[n_1 - n_2] = r^*_{yx}[-\ell] = \mathbb{E}\left[x[n]\; y^*[n-l]\right]$$
$$\gamma_{xy}[\ell] = \gamma_{xy}[n_1 - n_2] = \gamma^*_{yx}[-\ell] = r_{xy}[\ell] - \mu_x\, \mu^*_y$$

Joint-Ergodicity applies to two ergodic processes, $x[n]$ and $y[n]$, whose ensemble cross-correlation can be estimated from a time-average:

$$\langle x[n]\; y^*[n-l]\rangle = \mathbb{E}\left[x[n]\; y^*[n-l]\right]$$

Correlation Matrices

Let an M-dimensional random vector $\mathbf{X}[n, \zeta] \equiv \mathbf{X}[n]$ be derived from the random process $x[n]$ as follows:

$$\mathbf{X}[n] \triangleq \begin{bmatrix} x[n] & x[n-1] & \cdots & x[n-M+1] \end{bmatrix}^T$$

Correlation Matrices

Let an M-dimensional random vector $\mathbf{X}[n, \zeta] \equiv \mathbf{X}[n]$ be derived from the random process $x[n]$ as follows:

$$\mathbf{X}[n] \triangleq \begin{bmatrix} x[n] & x[n-1] & \cdots & x[n-M+1] \end{bmatrix}^T$$

Then its mean is given by an M-vector

$$\boldsymbol{\mu}_{\mathbf{X}}[n] \triangleq \begin{bmatrix} \mu_x[n] & \mu_x[n-1] & \cdots & \mu_x[n-M+1] \end{bmatrix}^T$$

Correlation Matrices

Let an M-dimensional random vector $\mathbf{X}[n,\zeta] \equiv \mathbf{X}[n]$ be derived from the random process $x[n]$ as follows:

$$\mathbf{X}[n] \triangleq \begin{bmatrix} x[n] & x[n-1] & \cdots & x[n-M+1] \end{bmatrix}^T$$

Then its mean is given by an M-vector

$$\boldsymbol{\mu}_{\mathbf{X}}[n] \triangleq \begin{bmatrix} \mu_x[n] & \mu_x[n-1] & \cdots & \mu_x[n-M+1] \end{bmatrix}^T$$

and the $M \times M$ correlation matrix is given by:

$$\mathbf{R}_{\mathbf{X}}[n] \triangleq \begin{bmatrix} r_{xx}[n,n] & \cdots & r_{xx}[n,n-M+1] \\ \vdots & \ddots & \vdots \\ r_{xx}[n-M+1,n] & \cdots & r_{xx}[n-M+1,n-M+1] \end{bmatrix}$$

Correlation Matrices

For WSS processes, the correlation matrix has:

1. $\mathbf{R_X}[n]$ is a constant matrix $\mathbf{R_X}$;
2. $r_{xx}[n-i, n-j] = r_{xx}[j-i] = r_{xx}[\ell]$, $\ell = j - i$;
3. conjugate symmetry gives $r_{xx}[\ell] = r_{xx}^*[-\ell]$.

Correlation Matrices

For WSS processes, the correlation matrix has:

1. $\mathbf{R_X}[n]$ is a constant matrix $\mathbf{R_X}$;
2. $r_{xx}[n-i, n-j] = r_{xx}[j-i] = r_{xx}[\ell]\,,\ \ell = j-i$;
3. conjugate symmetry gives $r_{xx}[\ell] = r^*_{xx}[-\ell]$.

Hence, the matrix $\mathbf{R}_{xx}$ is given by:

$$\mathbf{R_X} \triangleq \begin{bmatrix} r_{xx}[0] & r_{xx}[1] & r_{xx}[2] & \cdots & r_{xx}[M-1] \\ r^*_{xx}[1] & r_{xx}[0] & r_{xx}[1] & \cdots & r_{xx}[M-2] \\ r^*_{xx}[2] & r^*_{xx}[1] & r_{xx}[0] & \cdots & r_{xx}[M-3] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r^*_{xx}[M-1] & r^*_{xx}[M-2] & r^*_{xx}[M-3] & \cdots & r_{xx}[0] \end{bmatrix}$$

Correlation Matrices

For WSS processes, the correlation matrix has:

1. $\mathbf{R_X}[n]$ is a constant matrix $\mathbf{R_X}$;

2. $r_{xx}[n-i, n-j] = r_{xx}[j-i] = r_{xx}[\ell]\,,\ \ell = j - i$;

3. conjugate symmetry gives $r_{xx}[\ell] = r_{xx}^*[-\ell]$.

Hence, the matrix $\mathbf{R}_{xx}$ is given by:

$$\mathbf{R_X} \triangleq \begin{bmatrix} r_{xx}[0] & r_{xx}[1] & r_{xx}[2] & \cdots & r_{xx}[M-1] \\ r_{xx}^*[1] & r_{xx}[0] & r_{xx}[1] & \cdots & r_{xx}[M-2] \\ r_{xx}^*[2] & r_{xx}^*[1] & r_{xx}[0] & \cdots & r_{xx}[M-3] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{xx}^*[M-1] & r_{xx}^*[M-2] & r_{xx}^*[M-3] & \cdots & r_{xx}[0] \end{bmatrix}$$

It can be seen that $\mathbf{R_X}$ is Hermitian and **Toeplitz**.

Correlation Matrices

Example (Correlation matrices). The correlation function for a certain random process $x[n]$ has the exponential form:

$$r_{xx}[\ell] = 4\,(-0.5)^{|\ell|}$$

$\bowtie$

Correlation Matrices

Example (Correlation matrices). The correlation function for a certain random process $x[n]$ has the exponential form:

$$r_{xx}[\ell] = 4\,(-0.5)^{|\ell|}$$

Hence, the correlation matrix for $N = 3$ is given by:

$$\mathbf{R_X} = \begin{bmatrix} r_{xx}[0] & r_{xx}[1] & r_{xx}[2] \\ r_{xx}^*[1] & r_{xx}[0] & r_{xx}[1] \\ r_{xx}^*[2] & r_{xx}^*[1] & r_{xx}^*[0] \end{bmatrix}$$

$$= \begin{bmatrix} 4(-0.5)^0 & 4(-0.5)^1 & 4(-0.5)^2 \\ 4(-0.5)^1 & 4(-0.5)^0 & 4(-0.5)^1 \\ 4(-0.5)^2 & 4(-0.5)^1 & 4(-0.5)^0 \end{bmatrix} = \begin{bmatrix} 4 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 4 \end{bmatrix} \bowtie$$

which is clearly Toeplitz.

Correlation Matrices

– End-of-Topic 52: **Joint Statistics and Correlation Matrices** –

Any Questions?

Markov Processes

A powerful model for a stochastic process known as a **Markov model** is introduced; such a process that satisfies this model is known as a **Markov process**.

- Quite simply, a Markov process is one in which the probability of any particular value in a sequence is dependent upon the preceding sample values.

Markov Processes

A powerful model for a stochastic process known as a **Markov model** is introduced; such a process that satisfies this model is known as a **Markov process**.

- Quite simply, a Markov process is one in which the probability of any particular value in a sequence is dependent upon the preceding sample values.
- The simplest kind of dependence arises when the probability of any sample depends only upon the value of the *immediately preceding* sample, and this is known as a **first-order Markov process**.

Markov Processes

A powerful model for a stochastic process known as a **Markov model** is introduced; such a process that satisfies this model is known as a **Markov process**.

- Quite simply, a Markov process is one in which the probability of any particular value in a sequence is dependent upon the preceding sample values.
- The simplest kind of dependence arises when the probability of any sample depends only upon the value of the *immediately preceding* sample, and this is known as a **first-order Markov process**.
- This simple process is a surprisingly good model for a number of practical signal processing, communications and control problems.

Markov Processes

As an example of a Markov process, consider the process generated by the difference equation

$$x[n] = -a\,x[n-1] + w[n]$$

- where a is a known constant;
- and $w[n]$ is a sequence of zero-mean i. i. d. Gaussian random variables with variance σ_W^2 density:

$$f_W\left(w[n]\right) = \frac{1}{\sqrt{2\pi\sigma_W^2}} \exp\left\{-\frac{w^2[n]}{2\sigma_W^2}\right\}$$

Markov Processes

As an example of a Markov process, consider the process generated by the difference equation

$$x[n] = -a\,x[n-1] + w[n]$$

- where a is a known constant;
- and $w[n]$ is a sequence of zero-mean i. i. d. Gaussian random variables with variance σ_W^2 density:

$$f_W\left(w[n]\right) = \frac{1}{\sqrt{2\pi\sigma_W^2}} \exp\left\{-\frac{w^2[n]}{2\sigma_W^2}\right\}$$

The conditional density of $x[n]$ given $x[n-1]$ is also Gaussian,

$$f_X\left(x[n] \mid x[n-1]\right) = \frac{1}{\sqrt{2\pi\sigma_W^2}} \exp\left\{-\frac{(x[n]+ax[n-1])^2}{2\sigma_W^2}\right\}$$

Markov Processes

Definition (Markov Process). A random process is a Pth-order Markov process if the distribution of $x[n]$, given the infinite past, depends only on the previous P samples $\{x[n-1], \dots, x[n-P]\}$; that is, if:

$$f_X(x[n] \mid x[n-1], x[n-2], \dots) = f_X(x[n] \mid x[n-1], \dots, x[n-P])$$

♢

Markov Processes

Definition (Markov Process). A random process is a Pth-order Markov process if the distribution of $x[n]$, given the infinite past, depends only on the previous P samples $\{x[n-1], \dots, x[n-P]\}$; that is, if:

$$f_X\left(x[n] \mid x[n-1], x[n-2], \dots\right) = f_X\left(x[n] \mid x[n-1], \dots, x[n-P]\right)$$

$\diamondsuit$

Example (First-order Markov). A first-order Markov process is where, given the infinite past, the current sample of a random process $x[n]$ depends only on the previous sample $x[n-1]$:

$$f_X\left(x[n] \mid x[n-1], x[n-2], \dots, x[0]\right) = f_X\left(x[n] \mid x[n-1]\right)$$

$\bowtie$

Markov Processes

Example (First-order Markov). A first-order Markov process is where, given the infinite past, the current sample of a random process $x[n]$ depends only on the previous sample $x[n-1]$:

$$f_X\left(x[n] \mid x[n-1], x[n-2], \dots, x[0]\right) = f_X\left(x[n] \mid x[n-1]\right)$$

- Using the probability chain rule, and defining $\mathbf{x} = \{x[n], x[n-1], \dots, x[0]\}$, the general joint-pdfof all samples is:

$$f_{\mathbf{X}}(\mathbf{x}) = f_X\left(x[n] \mid x[n-1], x[n-2], \dots, x[0]\right) \\ \times f_X\left(x[n-1] \mid x[n-2], x[n-3], \dots, x[0]\right) \cdots f_X\left(x[0]\right)$$

$\bowtie$

Markov Processes

Example (First-order Markov). A first-order Markov process is where, given the infinite past, the current sample of a random process $x[n]$ depends only on the previous sample $x[n-1]$:

$$f_X\left(x[n] \mid x[n-1], x[n-2], \dots, x[0]\right) = f_X\left(x[n] \mid x[n-1]\right)$$

- Using the probability chain rule, the general joint-pdf:

$$f_{\mathbf{X}}(\mathbf{x}) = f_X\left(x[n] \mid x[n-1], x[n-2], \dots, x[0]\right) \times f_X\left(x[n-1] \mid x[n-2], x[n-3], \dots, x[0]\right) \cdots f_X\left(x[0]\right)$$

This can be written as:

$$f_{\mathbf{X}}(\mathbf{x}) = f_X\left(x[0]\right) \prod_{k=1}^{n} f_X\left(x[k] \mid x[k-1], \dots, x[0]\right)$$

Markov Processes

Example (First-order Markov). A first-order Markov process is where, given the infinite past, the current sample of a random process $x[n]$ depends only on the previous sample $x[n-1]$:

$$f_X\left(x[n] \mid x[n-1], x[n-2], \dots, x[0]\right) = f_X\left(x[n] \mid x[n-1]\right)$$

This can be written as:

$$f_{\mathbf{X}}(\mathbf{x}) = f_X(x[0]) \prod_{k=1}^{n} f_X\left(x[k] \mid x[k-1], \dots, x[0]\right)$$

Hence, using the first-order Markov property, this simplifies to:

$$f_{\mathbf{X}}(\mathbf{x}) = f_X(x[0]) \prod_{k=1}^{n} f_X\left(x[k] \mid x[k-1]\right)$$

$\bowtie$

Markov Processes

Example (First-order Markov). A first-order Markov process is where, given the infinite past, the current sample of a random process $x[n]$ depends only on the previous sample $x[n-1]$:

$$f_X\left(x[n] \mid x[n-1], x[n-2], \dots, x[0]\right) = f_X\left(x[n] \mid x[n-1]\right)$$

Hence, using the first-order Markov property, this simplifies to:

$$f_{\mathbf{X}}(\mathbf{x}) = f_X(x[0]) \prod_{k=1}^{n} f_X\left(x[k] \mid x[k-1]\right)$$

This allows us to substitute, for example, the Gaussian:

$$f_{\mathbf{X}}(\mathbf{x}) = f_X(x[0]) \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi\sigma_W^2}} \exp\left\{-\frac{(x[n] + a\,x[n-1])^2}{2\sigma_W^2}\right\} \quad \bowtie$$

Markov Processes

Finally, it is noted that if $x[n]$ takes on a countable (discrete) set of values, a Markov random process is called a **Markov chain**.

Markov Processes

– End-of-Topic 53: **Brief Introduction to Markov Processes** –

Any Questions?

Lecture Slideset 3
Power Spectral Density

Introduction

Frequency- and transform-domain methods are very powerful tools for the analysis of deterministic sequences. It seems natural to extend these techniques to analysis stationary **random processes**.

Introduction

Frequency- and transform-domain methods are very powerful tools for the analysis of deterministic sequences. It seems natural to extend these techniques to analysis stationary **random processes**.

So far in this course, **stationary stochastic process**es have been considered in the time-domain through the use of the **ACS**.

Introduction

Frequency- and transform-domain methods are very powerful tools for the analysis of deterministic sequences. It seems natural to extend these techniques to analysis stationary **random processes**.

So far in this course, **stationary stochastic process**es have been considered in the time-domain through the use of the **ACS**.

- Since the ACS for a stationary process is a function of a single-discrete time process, then the question arises as to what the discrete-time Fourier transform (DTFT) corresponds to.

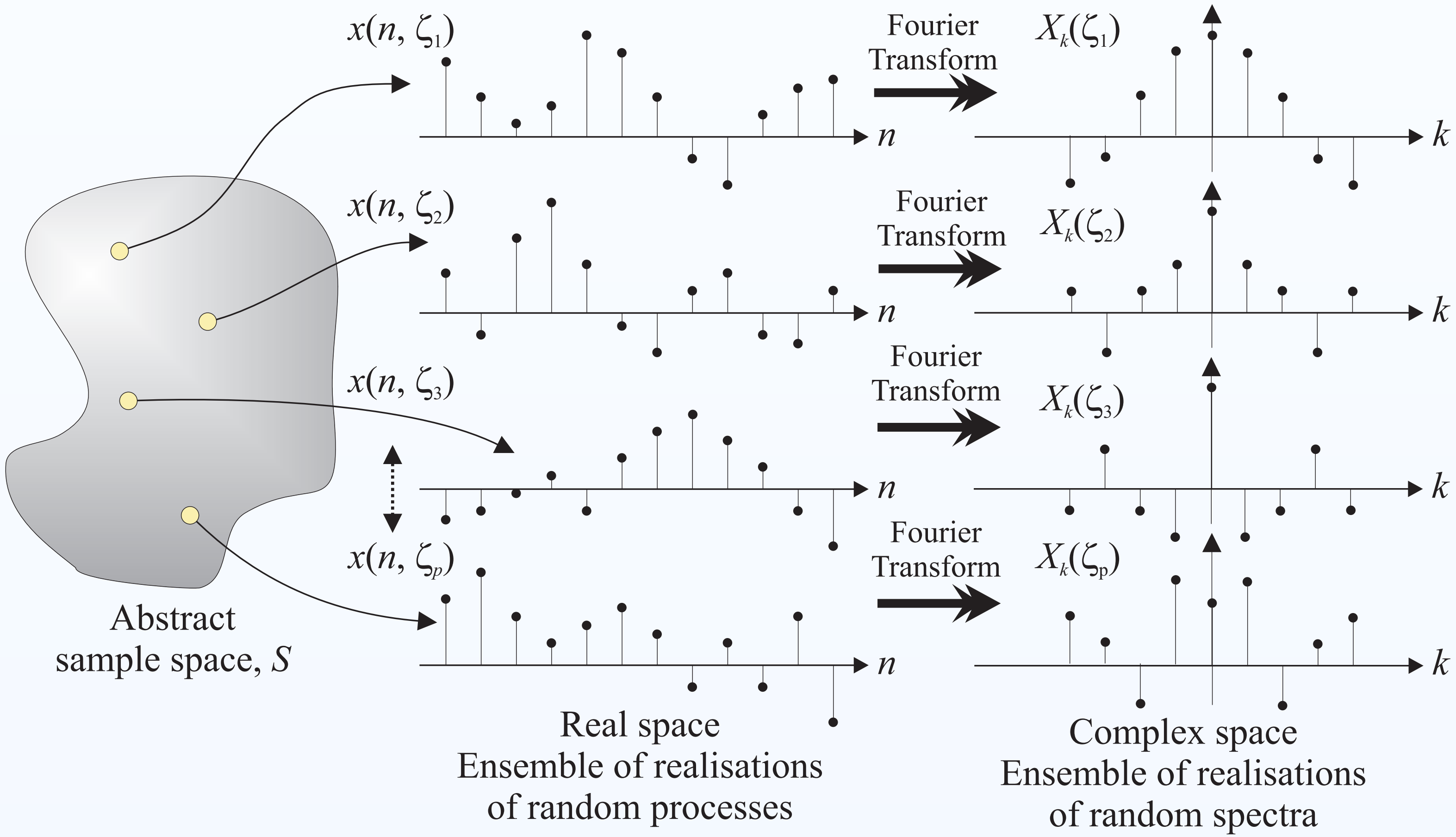
Introduction

Frequency- and transform-domain methods are very powerful tools for the analysis of deterministic sequences. It seems natural to extend these techniques to analysis stationary **random processes**.

So far in this course, **stationary stochastic process**es have been considered in the time-domain through the use of the **ACS**.

- Since the ACS for a stationary process is a function of a single-discrete time process, then the question arises as to what the DTFT corresponds to.
- It turns out to be known as the **power spectral density (PSD)** of a stationary random process, and the PSD is an extremely powerful and conceptually appealing tool in statistical signal processing.

Introduction



A graphical respresentation of random spectra.

Introduction

In signal theory for deterministic signals, spectra are used to represent a function as a superposition of exponential functions. For random signals, the notion of a spectrum has two interpretations:

Introduction

In signal theory for deterministic signals, spectra are used to represent a function as a superposition of exponential functions. For random signals, the notion of a spectrum has two interpretations:

Transform of averages The first involves transform of averages (or moments). As will be seen, this will be the Fourier transform of the autocorrelation function.

Introduction

In signal theory for deterministic signals, spectra are used to represent a function as a superposition of exponential functions. For random signals, the notion of a spectrum has two interpretations:

Transform of averages The first involves transform of averages (or moments). As will be seen, this will be the Fourier transform of the autocorrelation function.

Stochastic decomposition The second interpretation represents a stochastic process as a superposition of exponentials, where the coefficients are themselves random variables. Hence, $x[n]$ can be represented as:

$$x[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X\left(e^{j\omega T}\right) e^{j\omega n}\, d\omega, \quad n \in \mathbb{R}$$

where $X(e^{j\omega})$ is a random variable for a given value of ω.

Motivating the power spectral density

- It is important to appreciate that most realisations of stationary random signals, $x[n, \zeta]$, do not have finite energy, as they usually don't decay away as $n \to \pm\infty$.

Motivating the power spectral density

- It is important to appreciate that most realisations of stationary random signals, $x[n, \zeta]$, do not have finite energy, as they usually don't decay away as $n \to \pm\infty$.
- This is because the statistics as $n \to \pm\infty$ are the same as the statistics at any other time.

Motivating the power spectral density

- It is important to appreciate that most realisations of stationary random signals, $x[n, \zeta]$, do not have finite energy, as they usually don't decay away as $n \to \pm\infty$.
- This is because the statistics as $n \to \pm\infty$ are the same as the statistics at any other time.
- Therefore, technically, these realisations do not possess a corresponding DTFT, and hence it is not possible simply to take the DTFT of the random signal without further addressing these technicalities.

Motivating the power spectral density

- It is important to appreciate that most realisations of stationary random signals, $x[n, \zeta]$, do not have finite energy, as they usually don't decay away as $n \to \pm\infty$.
- This is because the statistics as $n \to \pm\infty$ are the same as the statistics at any other time.
- Therefore, technically, these realisations do not possess a corresponding DTFT, and hence it is not possible simply to take the DTFT of the random signal without further addressing these technicalities.
- Moreover, noting that a random signal is actually an ensemble of realisations, each realisation occuring with a different probability, it raises the question of what does it mean to take the DTFT of a random process directly?

Informal Motivation

- Assume for the moment that the DTFT of a realisation from a stationary random process does in fact exist, by ignoring any issues with convergence of the sequence.

Informal Motivation

- Assume for the moment that the DTFT of a realisation from a stationary random process does in fact exist, by ignoring any issues with convergence of the sequence.
- If a particular realisation is denoted by $x[n, \zeta]$, then suppose the corresponding DTFT is denoted by:

$$X_\zeta\left(e^{j\omega T}\right) = \sum_{n=-\infty}^{\infty} x[n,\zeta]\, e^{-j\omega n}$$

where $|\omega| < \pi$ is the normalised frequency.

Informal Motivation

- Assume for the moment that the DTFT of a realisation from a stationary random process does in fact exist, by ignoring any issues with convergence of the sequence.
- If a particular realisation is denoted by $x[n, \zeta]$, then suppose the corresponding DTFT is denoted by:

$$X_\zeta\left(e^{j\omega T}\right) = \sum_{n=-\infty}^{\infty} x[n,\zeta]\, e^{-j\omega n}$$

where $|\omega| < \pi$ is the normalised frequency.

- The collection of different DTFTs forms an ensemble of frequency-domain realisations.

Informal Motivation

- Assume for the moment that the DTFT of a realisation from a stationary random process does in fact exist, by ignoring any issues with convergence of the sequence.
- If a particular realisation is denoted by $x[n, \zeta]$, then suppose the corresponding DTFT is denoted by:

$$X_\zeta\left(e^{j\omega T}\right) = \sum_{n=-\infty}^{\infty} x[n, \zeta]\, e^{-j\omega n}$$

where $|\omega| < \pi$ is the normalised frequency.

- The collection of different DTFTs forms an ensemble of frequency-domain realisations.
- As this spectrum is continuous, the second-order ACF is a seemingly important statistic to consider, representing the correlation between two frequencies at ω_1 and ω_2, say.

Informal Motivation

Hence, consider forming:

$$R_{XX}(\omega_1, \omega_2) = \mathbb{E}\left[X_\zeta\left(e^{j\omega_1}\right) X_\zeta^*\left(e^{j\omega_2}\right)\right]$$

Substituting the DTFT expression, and reorganising:

Informal Motivation

Hence, consider forming:

$$R_{XX}(\omega_1, \omega_2) = \mathbb{E}\left[X_\zeta\left(e^{j\omega_1}\right) X_\zeta^*\left(e^{j\omega_2}\right)\right]$$

Substituting the DTFT expression, and reorganising:

$$R_{XX}(\omega_1, \omega_2) = \mathbb{E}\left[\sum_{n=-\infty}^{\infty} x[n, \zeta]\, e^{-j\omega_1 n} \sum_{m=-\infty}^{\infty} x^*[m, \zeta]\, e^{j\omega_2 m}\right]$$

Informal Motivation

Hence, consider forming:

$$R_{XX}(\omega_1, \omega_2) = \mathbb{E}\left[X_\zeta\left(e^{j\omega_1}\right) X_\zeta^*\left(e^{j\omega_2}\right)\right]$$

Substituting the DTFT expression, and reorganising:

$$R_{XX}(\omega_1, \omega_2) = \mathbb{E}\left[\sum_{n=-\infty}^{\infty} x[n,\zeta]\, e^{-j\omega_1 n} \sum_{m=-\infty}^{\infty} x^*[m,\zeta]\, e^{j\omega_2 m}\right]$$

$$= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \mathbb{E}\left[x[n,\zeta]\, x^*[m,\zeta]\right] e^{-j(\omega_1 n - \omega_2 m)}$$

Informal Motivation

Hence, consider forming:

$$R_{XX}(\omega_1, \omega_2) = \mathbb{E}\left[X_\zeta\left(e^{j\omega_1}\right) X_\zeta^*\left(e^{j\omega_2}\right)\right]$$

Substituting the DTFT expression, and reorganising:

$$\begin{aligned} R_{XX}(\omega_1, \omega_2) &= \mathbb{E}\left[\sum_{n=-\infty}^{\infty} x[n,\zeta]\, e^{-j\omega_1 n} \sum_{m=-\infty}^{\infty} x^*[m,\zeta]\, e^{j\omega_2 m}\right] \\ &= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \mathbb{E}\left[x[n,\zeta]\, x^*[m,\zeta]\right] e^{-j(\omega_1 n - \omega_2 m)} \\ &= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} r_{xx}[n,m]\, e^{-j(\omega_1 n - \omega_2 m)} \end{aligned}$$

Informal Motivation

Hence, consider forming:

$$R_{XX}(\omega_1, \omega_2) = \mathbb{E}\left[X_\zeta\left(e^{j\omega_1}\right) X_\zeta^*\left(e^{j\omega_2}\right)\right]$$

Substituting the DTFT expression, and reorganising:

$$\begin{aligned} R_{XX}(\omega_1, \omega_2) &= \mathbb{E}\left[\sum_{n=-\infty}^{\infty} x[n,\zeta]\, e^{-j\omega_1 n} \sum_{m=-\infty}^{\infty} x^*[m,\zeta]\, e^{j\omega_2 m}\right] \\ &= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \mathbb{E}\left[x[n,\zeta]\, x^*[m,\zeta]\right] e^{-j(\omega_1 n - \omega_2 m)} \\ &= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} r_{xx}[n,m]\, e^{-j(\omega_1 n - \omega_2 m)} \end{aligned}$$

It can be seen that it is indicative of some kind of Fourier transform of the corresponding time-domain correlation.

Informal Motivation

Hence, consider forming:

$$R_{XX}(\omega_1, \omega_2) = \mathbb{E}\left[X_\zeta\left(e^{j\omega_1}\right) X_\zeta^*\left(e^{j\omega_2}\right)\right]$$

Substituting the DTFT expression, and reorganising:

$$R_{XX}(\omega_1, \omega_2) = \mathbb{E}\left[\sum_{n=-\infty}^{\infty} x[n,\zeta]\, e^{-j\omega_1 n} \sum_{m=-\infty}^{\infty} x^*[m,\zeta]\, e^{j\omega_2 m}\right]$$

$$= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} r_{xx}[n,m]\, e^{-j(\omega_1 n - \omega_2 m)}$$

- Indeed, as $x[n,\zeta]$ is stationary, then let $r_{xx}[n,m] = r_{xx}[n-m]$.

Informal Motivation

Hence, consider forming:

$$R_{XX}(\omega_1, \omega_2) = \mathbb{E}\left[X_\zeta\left(e^{j\omega_1}\right) X_\zeta^*\left(e^{j\omega_2}\right)\right]$$

Substituting the DTFT expression, and reorganising:

$$R_{XX}(\omega_1, \omega_2) = \mathbb{E}\left[\sum_{n=-\infty}^{\infty} x[n,\zeta]\, e^{-j\omega_1 n} \sum_{m=-\infty}^{\infty} x^*[m,\zeta]\, e^{j\omega_2 m}\right]$$

$$= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} r_{xx}[n,m]\, e^{-j(\omega_1 n - \omega_2 m)}$$

- Indeed, as $x[n,\zeta]$ is stationary, then let $r_{xx}[n,m] = r_{xx}[n-m]$.
- Consider finding the second-moment or power at a given frequency, so setting $\omega = \omega_1 = \omega_2$, and $\ell = n - m$.

Informal Motivation

Hence, consider forming:

$$R_{XX}(\omega_1, \omega_2) = \mathbb{E}\left[X_\zeta\left(e^{j\omega_1}\right) X_\zeta^*\left(e^{j\omega_2}\right)\right]$$

Substituting the DTFT expression, and reorganising:

$$R_{XX}(\omega_1, \omega_2) = \mathbb{E}\left[\sum_{n=-\infty}^{\infty} x[n,\zeta]\, e^{-j\omega_1 n} \sum_{m=-\infty}^{\infty} x^*[m,\zeta]\, e^{j\omega_2 m}\right]$$

- Then, it follows that:

$$R_{XX}(\omega) = \sum_{n=-\infty}^{\infty} \sum_{\ell=-\infty}^{\infty} r_{xx}[\ell]\, e^{-j\omega\ell} = \sum_{n=-\infty}^{\infty} \mathcal{F}\left(r_{xx}[\ell]\right)$$

Informal Motivation

Hence, consider forming:

$$R_{XX}(\omega_1, \omega_2) = \mathbb{E}\left[X_\zeta\left(e^{j\omega_1}\right) X_\zeta^*\left(e^{j\omega_2}\right)\right]$$

- Then, it follows that:

$$R_{XX}(\omega) = \sum_{n=-\infty}^{\infty} \sum_{\ell=-\infty}^{\infty} r_{xx}[\ell]\, e^{-j\omega\ell} = \sum_{n=-\infty}^{\infty} \mathcal{F}\left(r_{xx}[\ell]\right)$$

- The additional summation results from the fact the realisations of the process do not have finite-energy, and the mathematical treatment somewhat informal.
- However, it clearly indicates that the power at each frequency can be found from the Fourier transform of the ACS, and is therefore the PSD.

Formal Statistical Derivation

Consider the random variable, $X\left(e^{j\omega T}\right)$, resulting from the DTFT of a random signal, $x[n]$:

$$X\left(e^{j\omega T}\right) = \sum_{n=-\infty}^{\infty} x[n]\, e^{-j\omega n}$$

Formal Statistical Derivation

Consider the random variable, $X\left(e^{j\omega T}\right)$:

$$X\left(e^{j\omega T}\right) = \sum_{n=-\infty}^{\infty} x[n]\, e^{-j\omega n}$$

Consider the **total power** in $X\left(e^{j\omega T}\right)$:

$$P_{XX}\left(e^{j\omega T}\right) = \mathbb{E}\left[\left|X\left(e^{j\omega T}\right)\right|^2\right]$$

Formal Statistical Derivation

Consider the random variable, $X\left(e^{j\omega T}\right)$:

$$X\left(e^{j\omega T}\right) = \sum_{n=-\infty}^{\infty} x[n]\, e^{-j\omega n}$$

Consider the **total power** in $X\left(e^{j\omega T}\right)$:

$$P_{XX}\left(e^{j\omega T}\right) = \mathbb{E}\left[\left|X\left(e^{j\omega T}\right)\right|^2\right]$$

Since this expression will diverge, so consider:

$$P_{XX}\left(e^{j\omega T}\right) = \lim_{N\to\infty} \frac{1}{2N+1} \mathbb{E}\left[\left|X_N\left(e^{j\omega}\right)\right|^2\right]$$

Formal Statistical Derivation

Consider the random variable, $X\left(e^{j\omega T}\right)$:

$$X\left(e^{j\omega T}\right) = \sum_{n=-\infty}^{\infty} x[n]\, e^{-j\omega n}$$

Consider the **total power** in $X\left(e^{j\omega T}\right)$:

$$P_{XX}\left(e^{j\omega T}\right) = \mathbb{E}\left[\left|X\left(e^{j\omega T}\right)\right|^2\right]$$

Since this expression will diverge, so consider:

$$P_{XX}\left(e^{j\omega T}\right) = \lim_{N\to\infty} \frac{1}{2N+1} \mathbb{E}\left[\left|X_N\left(e^{j\omega}\right)\right|^2\right]$$

where $X_N\left(e^{j\omega}\right)$ is a **windowed** version of $x[n]$:

$$X_N\left(e^{j\omega T}\right) \triangleq \sum_{n=-N}^{N} x[n]\, e^{-j\omega n}$$

Formal Statistical Derivation

Since this expression will diverge, so consider:

$$P_{XX}\left(e^{j\omega T}\right) = \lim_{N\to\infty} \frac{1}{2N+1} \mathbb{E}\left[\left|X_N\left(e^{j\omega}\right)\right|^2\right]$$

Then, substituting and rearranging gives:

$$P_{XX}\left(e^{j\omega T}\right) = \lim_{N\to\infty} \frac{1}{2N+1} \mathbb{E}\left[\sum_{n=-N}^{N} x[n]\, e^{-j\omega n} \sum_{m=-N}^{N} x^*[m]\, e^{j\omega m}\right]$$

$$= \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} \sum_{m=-N}^{N} \mathbb{E}\left[x[n]\, x^*[m]\right] e^{-j\omega(n-m)}$$

Formal Statistical Derivation

Since this expression will diverge, so consider:

$$P_{XX}\left(e^{j\omega T}\right) = \lim_{N\to\infty} \frac{1}{2N+1} \mathbb{E}\left[\left|X_N\left(e^{j\omega}\right)\right|^2\right]$$

Then, substituting and rearranging gives:

$$P_{XX}\left(e^{j\omega T}\right) = \lim_{N\to\infty} \frac{1}{2N+1} \mathbb{E}\left[\sum_{n=-N}^{N} x[n]\, e^{-j\omega n} \sum_{m=-N}^{N} x^*[m]\, e^{j\omega m}\right]$$

It can be shown this expression simplifies to DTFT of the ACS.

$$P_{XX}\left(e^{j\omega}\right) = \sum_{\ell=-\infty}^{\infty} r_{xx}[\ell]\, e^{-j\omega\ell}$$

Formal Statistical Derivation

Since this expression will diverge, so consider:

$$P_{XX}\left(e^{j\omega T}\right) = \lim_{N\to\infty} \frac{1}{2N+1} \mathbb{E}\left[\left|X_N\left(e^{j\omega}\right)\right|^2\right]$$

It can be shown this expression simplifies to DTFT of the ACS.

$$P_{XX}\left(e^{j\omega}\right) = \sum_{\ell=-\infty}^{\infty} r_{xx}[\ell]\, e^{-j\omega\ell}$$

- Hence, $P_{XX}\left(e^{j\omega T}\right)$ can be viewed as the average power, or energy, of the Fourier transform of a random process at frequency ω.
- Clearly, this gives an indication of whether, *on average*, there are dominant frequencies present in the realisations of $x[n]$.

Formal Statistical Derivation

– End-of-Topic 54: **Introduction to the concept of the PSD** –

Any Questions?

The power spectral density

The discrete-time Fourier transform of the autocorrelation sequence of a stationary stochastic process $x[n, \zeta]$ is known as the **power spectral density (PSD)**, is denoted by $P_{xx}(e^{j\omega})$, and is given by:

$$P_{xx}(e^{j\omega}) = \sum_{\ell \in \mathbb{Z}} r_{xx}[\ell] \; e^{-j\omega\ell}$$

where ω is frequency in radians per sample.

The power spectral density

The discrete-time Fourier transform of the autocorrelation sequence of a stationary stochastic process $x[n, \zeta]$ is known as the **power spectral density (PSD)**, is denoted by $P_{xx}(e^{j\omega})$, and is given by:

$$P_{xx}(e^{j\omega}) = \sum_{\ell \in \mathbb{Z}} r_{xx}[\ell]\, e^{-j\omega\ell}$$

where ω is frequency in radians per sample.

The autocorrelation sequence, $r_{xx}[\ell]$, can be recovered from the **PSD** by using the inverse-**DTFT**:

$$r_{xx}[\ell] = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_{xx}(e^{j\omega})\, e^{j\omega\ell}\, d\omega, \quad \ell \in \mathbb{Z}$$

Properties of the power spectral density

- $P_{xx}(e^{j\omega}) : \omega \to \mathbb{R}^+$; in otherwords, the PSD is real valued, and nonnegative definite. i.e.

$$P_{xx}\left(e^{j\omega T}\right) \geq 0$$

Properties of the power spectral density

- $P_{xx}(e^{j\omega}) : \omega \to \mathbb{R}^+$; in otherwords, the PSD is real valued, and nonnegative definite. i.e.

$$P_{xx}\left(e^{j\omega T}\right) \geq 0$$

- $P_{xx}(e^{j\omega}) = P_{xx}(e^{j(\omega+2n\pi)})$; in otherwords, the PSD is periodic with period 2π.

Properties of the power spectral density

- $P_{xx}(e^{j\omega}) : \omega \to \mathbb{R}^+$; in otherwords, the PSD is real valued, and nonnegative definite. i.e.

$$P_{xx}\left(e^{j\omega T}\right) \geq 0$$

- $P_{xx}(e^{j\omega}) = P_{xx}(e^{j(\omega+2n\pi)})$; periodic with period 2π.
- If $x[n]$ is real-valued, then:

Properties of the power spectral density

- $P_{xx}(e^{j\omega}) : \omega \rightarrow \mathbb{R}^+$; in otherwords, the PSD is real valued, and nonnegative definite. i.e.

$$P_{xx}\left(e^{j\omega T}\right) \geq 0$$

- $P_{xx}(e^{j\omega}) = P_{xx}(e^{j(\omega+2n\pi)})$; periodic with period 2π.
- If $x[n]$ is real-valued, then:
  - $r_{xx}[\ell]$ is real and even;

Properties of the power spectral density

- $P_{xx}(e^{j\omega}) : \omega \to \mathbb{R}^+$; in otherwords, the PSD is real valued, and nonnegative definite. i.e.

$$P_{xx}\left(e^{j\omega T}\right) \geq 0$$

- $P_{xx}(e^{j\omega}) = P_{xx}(e^{j(\omega+2n\pi)})$; periodic with period 2π.
- If $x[n]$ is real-valued, then:
  - $r_{xx}[\ell]$ is real and even;
  - $P_{xx}(e^{j\omega}) = P_{xx}(e^{-j\omega})$ is an even function of ω.

Properties of the power spectral density

- $P_{xx}(e^{j\omega}) : \omega \to \mathbb{R}^+$; in otherwords, the PSD is real valued, and nonnegative definite. i.e.

$$P_{xx}\left(e^{j\omega T}\right) \geq 0$$

- $P_{xx}(e^{j\omega}) = P_{xx}(e^{j(\omega+2n\pi)})$; periodic with period 2π.
- If $x[n]$ is real-valued, then:
  - $r_{xx}[\ell]$ is real and even;
  - $P_{xx}(e^{j\omega}) = P_{xx}(e^{-j\omega})$ is an even function of ω.
- The area under $P_{xx}(e^{j\omega})$ is nonnegative and is equal to the average power of $x[n]$. Hence:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} P_{xx}(e^{j\omega})\, d\omega = r_{xx}[0] = \mathbb{E}\left[|x[n]|^2\right] \geq 0$$

Properties of the power spectral density

Example ([Manolakis:2001, Example 3.3.4, Page 109]). Determine the PSD of a zero-mean WSS process $x[n]$ with autocorrelation sequence $r_{xx}[\ell] = a^{|\ell|}$, $-1 < a < 1$.

Properties of the power spectral density

Example ([Manolakis:2001, Example 3.3.4, Page 109]). Determine the PSD of a zero-mean WSS process $x[n]$ with autocorrelation sequence $r_{xx}[\ell] = a^{|\ell|}$, $-1 < a < 1$.

SOLUTION. Using the definition of the PSD directly, then:

$$P_{xx}(e^{j\omega}) = \sum_{\ell \in \mathbb{Z}} r_{xx}[\ell]\, e^{-j\omega\ell}$$

□

Properties of the power spectral density

Example ([Manolakis:2001, Example 3.3.4, Page 109]). Determine the PSD of a zero-mean WSS process $x[n]$ with autocorrelation sequence $r_{xx}[\ell] = a^{|\ell|}$, $-1 < a < 1$.

SOLUTION. Using the definition of the PSD directly, then:

$$P_{xx}(e^{j\omega}) = \sum_{\ell \in \mathbb{Z}} r_{xx}[\ell] \, e^{-j\omega\ell}$$

$$= \sum_{\ell \in \mathbb{Z}} a^{|\ell|} \, e^{-j\omega\ell}$$

□

Properties of the power spectral density

Example ([Manolakis:2001, Example 3.3.4, Page 109]). Determine the PSD of a zero-mean WSS process $x[n]$ with autocorrelation sequence $r_{xx}[\ell] = a^{|\ell|}$, $-1 < a < 1$.

SOLUTION. Using the definition of the PSD directly, then:

$$
\begin{aligned}
P_{xx}(e^{j\omega}) &= \sum_{\ell \in \mathbb{Z}} r_{xx}[\ell] \; e^{-j\omega\ell} \\
&= \sum_{\ell \in \mathbb{Z}} a^{|\ell|} \; e^{-j\omega\ell} \\
&= \sum_{\ell=0}^{\infty} \left(a \, e^{-j\omega}\right)^{\ell} + \sum_{\ell=0}^{\infty} \left(a \, e^{j\omega}\right)^{\ell} - 1
\end{aligned}
$$

□

Properties of the power spectral density

Example ([Manolakis:2001, Example 3.3.4, Page 109]). Determine the PSD of a zero-mean WSS process $x[n]$ with autocorrelation sequence $r_{xx}[\ell] = a^{|\ell|}$, $-1 < a < 1$.

SOLUTION. Hence, by using the expressions for geometric series, the PSD can be written as:

$$\begin{aligned} P_{xx}(e^{j\omega}) &= \frac{1}{1 - a\, e^{-j\omega}} + \frac{1}{1 - a\, e^{j\omega}} - 1 \\ &= \frac{1 - a^2}{1 - 2a\cos\omega + a^2} \end{aligned} \qquad \square$$

which is a real-valued, even, and nonnegative function of ω.

General form of the PSD

A process, $x[n]$, and therefore $r_{xx}[\ell]$, can always be decomposed into a zero-mean aperiodic component, $r_{xx}^{(a)}[\ell]$, and a non-zero-mean periodic component, $r_{xx}^{(p)}[\ell]$:

$$r_{xx}[\ell] = r_{xx}^{(a)}[\ell] + r_{xx}^{(p)}[\ell]$$

General form of the PSD

A process, $x[n]$, and therefore $r_{xx}[\ell]$, can always be decomposed into a zero-mean aperiodic component, $r_{xx}^{(a)}[\ell]$, and a non-zero-mean periodic component, $r_{xx}^{(p)}[\ell]$:

$$r_{xx}[\ell] = r_{xx}^{(a)}[\ell] + r_{xx}^{(p)}[\ell]$$

Theorem (PSD of a non-zero-mean process with periodic component). The most general definition of the PSD for a non-zero-mean stochastic process with a periodic component is

$$P_{xx}(e^{j\omega}) = P_{xx}^{(a)}(e^{j\omega}) + \frac{2\pi}{K} \sum_{k \in \mathcal{K}} P_{xx}^{(p)}(k)\, \delta\left(\omega - \omega_k\right) \qquad \diamondsuit$$

$P_{xx}^{(a)}(e^{j\omega})$ is the DTFT of $r_{xx}^{(a)}[\ell]$, while $P_{xx}^{(p)}(k)$ are the discrete Fourier transform (DFT) coefficients for $r_{xx}^{(p)}[\ell]$.

General form of the PSD

Example ([Manolakis:2001, Harmonic Processes, Page 110-111]). Determine the PSD of the **harmonic process** defined by:

$$x[n] = \sum_{k=1}^{M} A_k \cos(\omega_k n + \phi_k), \quad \omega_k \neq 0$$

$\bowtie$

General form of the PSD

Example ([Manolakis:2001, Harmonic Processes, Page 110-111]). Determine the PSD of the **harmonic process** defined by:

$$x[n] = \sum_{k=1}^{M} A_k \cos(\omega_k n + \phi_k), \quad \omega_k \neq 0$$

SOLUTION. $x[n]$ is a zero-mean stationary process, and ACS:

$$r_{xx}[\ell] = \frac{1}{2} \sum_{k=1}^{M} |A_k|^2 \cos \omega_k \ell, \quad -\infty < \ell < \infty$$

□

General form of the PSD

Example ([Manolakis:2001, Harmonic Processes, Page 110-111]).
Determine the PSD of the **harmonic process** defined by:

$$x[n] = \sum_{k=1}^{M} A_k \cos(\omega_k n + \phi_k), \quad \omega_k \neq 0$$

SOLUTION. $x[n]$ is a zero-mean stationary process, and ACS:

$$r_{xx}[\ell] = \frac{1}{2} \sum_{k=1}^{M} |A_k|^2 \cos \omega_k \ell, \quad -\infty < \ell < \infty$$

Hence, the ACS can be written as:

$$r_{xx}[\ell] = \sum_{k=-M}^{M} \frac{|A_k|^2}{4} e^{j\omega_k \ell}, \quad -\infty < \ell < \infty \qquad \square$$

where: $A_0 = 0$, $A_k = A_{-k}$, and $\omega_{-k} = -\omega_k$.

General form of the PSD

Example ([Manolakis:2001, Harmonic Processes, Page 110-111]). Determine the PSD of the **harmonic process** defined by:

$$x[n] = \sum_{k=1}^{M} A_k \cos(\omega_k n + \phi_k), \quad \omega_k \neq 0$$

SOLUTION. Hence, the ACS can be written as:

$$r_{xx}[\ell] = \sum_{k=-M}^{M} \frac{|A_k|^2}{4} e^{j\omega_k \ell}, \quad -\infty < \ell < \infty$$

Hence, it directly follows

$$P_{xx}(e^{j\omega}) = 2\pi \sum_{k=-M}^{M} \frac{|A_k|^2}{4} \delta(\omega - \omega_k) = \frac{\pi}{2} \sum_{k=-M}^{M} |A_k|^2 \delta(\omega - \omega_k) \quad \square$$

General form of the PSD

– End-of-Topic 55: **Definition and examples of the PSD for WSS processes** –

Any Questions?

The cross-power spectral density

The cross-power spectral density (CPSD) of two jointly stationary stochastic processes, $x[n]$ and $y[n]$, provides a description of their statistical relations in the frequency domain.

- It is defined, naturally, as the DTFT of the cross-correlation, $r_{xy}[\ell] \triangleq \mathbb{E}\left[x[n]\ y^*[n-\ell]\right]$:

$$P_{xy}\left(e^{j\omega T}\right) = \mathcal{F}\{r_{xy}[\ell]\} = \sum_{\ell\in\mathbb{Z}} r_{xy}[\ell]\ e^{-j\omega\ell}$$

The cross-power spectral density

The cross-power spectral density (CPSD) of two jointly stationary stochastic processes, $x[n]$ and $y[n]$, provides a description of their statistical relations in the frequency domain.

- It is defined, naturally, as the DTFT of the cross-correlation, $r_{xy}[\ell] \triangleq \mathbb{E}\left[x[n]\; y^*[n-\ell]\right]$:

$$P_{xy}\left(e^{j\omega T}\right) = \mathcal{F}\{r_{xy}[\ell]\} = \sum_{\ell \in \mathbb{Z}} r_{xy}[\ell]\; e^{-j\omega\ell}$$

The cross-correlation $r_{xy}[\ell]$ can be recovered by using the inverse-DTFT:

$$r_{xy}[\ell] = \frac{1}{2\pi}\int_{-\pi}^{\pi} P_{xy}\left(e^{j\omega T}\right)\, e^{j\omega\ell}\, d\omega, \quad \ell \in \mathbb{R}$$

The cross-power spectral density

The cross-power spectral density (CPSD) of two jointly stationary stochastic processes, $x[n]$ and $y[n]$, provides a description of their statistical relations in the frequency domain.

- It is defined, naturally, as the DTFT of the cross-correlation, $r_{xy}[\ell] \triangleq \mathbb{E}\left[x[n]\; y^*[n-\ell]\right]$:

$$P_{xy}\left(e^{j\omega T}\right) = \mathcal{F}\{r_{xy}[\ell]\} = \sum_{\ell\in\mathbb{Z}} r_{xy}[\ell]\; e^{-j\omega\ell}$$

The cross-correlation $r_{xy}[\ell]$ can be recovered by using the inverse-DTFT:

$$r_{xy}[\ell] = \frac{1}{2\pi}\int_{-\pi}^{\pi} P_{xy}\left(e^{j\omega T}\right)\, e^{j\omega\ell}\, d\omega, \quad \ell\in\mathbb{R}$$

The cross-spectrum $P_{xy}\left(e^{j\omega T}\right)$ is, in general, a complex function of ω.

The cross-power spectral density

Some properties of the CPSD and related definitions include:

1. $P_{xy}\left(e^{j\omega T}\right)$ is periodic in ω with period 2π.

The cross-power spectral density

Some properties of the CPSD and related definitions include:

1. $P_{xy}\left(e^{j\omega T}\right)$ is periodic in ω with period 2π.

2. Since $r_{xy}[\ell] = r^*_{yx}[-\ell]$, then it follows:

$$P_{xy}\left(e^{j\omega T}\right) = P^*_{yx}\left(e^{j\omega T}\right)$$

3. If the process $x[n]$ is real, then $r_{xy}[\ell]$ is real, and:

$$P_{xy}(e^{j\omega}) = P^*_{xy}(e^{-j\omega})$$

The cross-power spectral density

Some properties of the CPSD and related definitions include:

1. $P_{xy}\left(e^{j\omega T}\right)$ is periodic in ω with period 2π.
2. Since $r_{xy}[\ell] = r^*_{yx}[-\ell]$, then it follows:

$$P_{xy}\left(e^{j\omega T}\right) = P^*_{yx}\left(e^{j\omega T}\right)$$

3. If the process $x[n]$ is real, then $r_{xy}[\ell]$ is real, and:

$$P_{xy}(e^{j\omega}) = P^*_{xy}(e^{-j\omega})$$

4. The **coherence function**, is given by:

$$\Gamma_{xy}(e^{j\omega}) \triangleq \frac{P_{xy}(e^{j\omega})}{\sqrt{P_{xx}(e^{j\omega})}\sqrt{P_{yy}(e^{j\omega})}}$$

Complex Spectral Density Functions

The second moment quantities that described a random process in the z-transform domain are known as the **complex spectral density** and **complex cross-spectral density** functions.

Complex Spectral Density Functions

The second moment quantities that described a random process in the z-transform domain are known as the **complex spectral density** and **complex cross-spectral density** functions.

Hence, $r_{xx}[\ell] \overset{z}{\rightleftharpoons} P_{xx}(z)$ and $r_{xy}[\ell] \overset{z}{\rightleftharpoons} P_{xy}(z)$, where:

$$P_{xx}(z) = \sum_{\ell \in \mathbb{Z}} r_{xx}[\ell]\, z^{-\ell}$$

$$P_{xy}(z) = \sum_{\ell \in \mathbb{Z}} r_{xy}[\ell]\, z^{-\ell}$$

Complex Spectral Density Functions

The second moment quantities that described a random process in the z-transform domain are known as the **complex spectral density** and **complex cross-spectral density** functions.

Hence, $r_{xx}[\ell] \overset{z}{\rightleftharpoons} P_{xx}(z)$ and $r_{xy}[\ell] \overset{z}{\rightleftharpoons} P_{xy}(z)$, where:

$$P_{xx}(z) = \sum_{\ell \in \mathbb{Z}} r_{xx}[\ell]\; z^{-\ell}$$

$$P_{xy}(z) = \sum_{\ell \in \mathbb{Z}} r_{xy}[\ell]\; z^{-\ell}$$

If the unit circle, defined by $z = e^{j\omega}$ is within the region of convergence of these summations, then:

$$P_{xx}(e^{j\omega}) = P_{xx}(z)\big|_{z=e^{j\omega}}$$

$$P_{xy}(e^{j\omega}) = P_{xy}(z)\big|_{z=e^{j\omega}}$$

Complex Spectral Density Functions

Example (Interleaved Example). Find the complex spectral-density of the sequence:

$$r[n] = \begin{cases} a^{\left|\frac{n}{2}\right|} & n \in \{0, \text{even}\} \\ 0 & \text{for } n \text{ odd} \end{cases}$$

$\bowtie$

Complex Spectral Density Functions

Example (Interleaved Example). Find complex spectral-density of :

$$r[n] = \begin{cases} a^{\left|\frac{n}{2}\right|} & n \in \{0, \text{even}\} \\ 0 & \text{for } n \text{ odd} \end{cases}$$

SOLUTION. Writing the z-transform:

$$P(z) = \sum_{\ell=-\infty}^{\infty} r[\ell]\, z^{-\ell}$$

□

Complex Spectral Density Functions

Example (Interleaved Example). Find complex spectral-density of :

$$r[n] = \begin{cases} a^{\left|\frac{n}{2}\right|} & n \in \{0, \text{even}\} \\ 0 & \text{for } n \text{ odd} \end{cases}$$

SOLUTION. Writing the z-transform:

$$P(z) = \sum_{\ell=-\infty}^{\infty} r[\ell]\, z^{-\ell}$$

$$= \underbrace{\sum_{\ell_o=-\infty}^{\infty} r[2\ell_o + 1]\, z^{-(2\ell_o+1)}}_{\text{Odd terms}} + \underbrace{\sum_{\ell_e=-\infty}^{\infty} r[2\ell_e]\, z^{-2\ell_e}}_{\text{Even terms}}$$

□

Complex Spectral Density Functions

Example (Interleaved Example). Find complex spectral-density of :

$$r[n] = \begin{cases} a^{\left|\frac{n}{2}\right|} & n \in \{0, \text{even}\} \\ 0 & \text{for } n \text{ odd} \end{cases}$$

SOLUTION. Writing the z-transform:

$$P(z) = \sum_{\ell=-\infty}^{\infty} r[\ell]\; z^{-\ell}$$

$$= \underbrace{\sum_{\ell_o=-\infty}^{\infty} r[2\ell_o + 1]\; z^{-(2\ell_o+1)}}_{\text{Odd terms}} + \underbrace{\sum_{\ell_e=-\infty}^{\infty} r[2\ell_e]\; z^{-2\ell_e}}_{\text{Even terms}}$$

$$= \sum_{\ell_e=-\infty}^{\infty} a^{\left|\frac{2\ell_e}{2}\right|}\; z^{-2\ell_e} = \sum_{\ell_e=-\infty}^{\infty} a^{|\ell_e|}\; z^{-2\ell_e}$$

Complex Spectral Density Functions

Example (Interleaved Example). Find complex spectral-density of :

$$r[n] = \begin{cases} a^{\left|\frac{n}{2}\right|} & n \in \{0, \text{even}\} \\ 0 & \text{for } n \text{ odd} \end{cases}$$

SOLUTION. Splitting this into two further summations, as previous done with an earlier example:

$$P(z) = \sum_{\ell_e=-\infty}^{0} a^{-\ell_e} z^{-2\ell_e} + \sum_{\ell_e=0}^{\infty} a^{\ell_e} z^{-2\ell_e} - 1$$

□

Complex Spectral Density Functions

Example (Interleaved Example). Find complex spectral-density of :

$$r[n] = \begin{cases} a^{\left|\frac{n}{2}\right|} & n \in \{0, \text{even}\} \\ 0 & \text{for } n \text{ odd} \end{cases}$$

SOLUTION. Splitting this into two further summations, as previous done with an earlier example:

$$P(z) = \sum_{\ell_e=-\infty}^{0} a^{-\ell_e} z^{-2\ell_e} + \sum_{\ell_e=0}^{\infty} a^{\ell_e} z^{-2\ell_e} - 1$$

$$= \sum_{\ell_e=0}^{\infty} (a\, z^2)^{\ell_e} + \sum_{\ell_e=0}^{\infty} \left(\frac{a}{z^2}\right)^{\ell_e} - 1$$

□

Complex Spectral Density Functions

Example (Interleaved Example). Find complex spectral-density of :

$$r[n] = \begin{cases} a^{\left|\frac{n}{2}\right|} & n \in \{0, \text{even}\} \\ 0 & \text{for } n \text{ odd} \end{cases}$$

SOLUTION. Splitting this into two further summations, as previous done with an earlier example:

$$P(z) = \sum_{\ell_e=-\infty}^{0} a^{-\ell_e} z^{-2\ell_e} + \sum_{\ell_e=0}^{\infty} a^{\ell_e} z^{-2\ell_e} - 1$$

$$= \sum_{\ell_e=0}^{\infty} \left(a z^2\right)^{\ell_e} + \sum_{\ell_e=0}^{\infty} \left(\frac{a}{z^2}\right)^{\ell_e} - 1 \qquad \square$$

Finally, applying the geometric progression formula $\sum_{\ell=0}^{\infty} r^{\ell} = \frac{1}{1-r}$ gives the desired result:

Complex Spectral Density Functions

Example (Interleaved Example). Find complex spectral-density of :

$$r[n] = \begin{cases} a^{\left|\frac{n}{2}\right|} & n \in \{0, \text{even}\} \\ 0 & \text{for } n \text{ odd} \end{cases}$$

SOLUTION. Finally, applying the geometric progression formula $\sum_{\ell=0}^{\infty} r^{\ell} = \frac{1}{1-r}$ gives the desired result:

$$P(z) = \frac{1}{1 - a\, z^2} + \frac{1}{1 - a\, z^{-2}} - 1$$

□

Complex Spectral Density Functions

Example (Interleaved Example). Find complex spectral-density of :

$$r[n] = \begin{cases} a^{\left|\frac{n}{2}\right|} & n \in \{0, \text{even}\} \\ 0 & \text{for } n \text{ odd} \end{cases}$$

SOLUTION. Finally, applying the geometric progression formula $\sum_{\ell=0}^{\infty} r^\ell = \frac{1}{1-r}$ gives the desired result:

$$P(z) = \frac{1}{1 - a\,z^2} + \frac{1}{1 - a\,z^{-2}} - 1$$
$$= \frac{1}{1 - a\,z^2} + \frac{a\,z^{-2}}{1 - a\,z^{-2}}$$

□

Complex Spectral Density Functions

Example (Interleaved Example). Find complex spectral-density of :

$$r[n] = \begin{cases} a^{\left|\frac{n}{2}\right|} & n \in \{0, \text{even}\} \\ 0 & \text{for } n \text{ odd} \end{cases}$$

SOLUTION. Finally, applying the geometric progression formula $\sum_{\ell=0}^{\infty} r^\ell = \frac{1}{1-r}$ gives the desired result:

$$P(z) = \frac{1}{1 - a\,z^2} + \frac{1}{1 - a\,z^{-2}} - 1$$
$$= \frac{1}{1 - a\,z^2} + \frac{a\,z^{-2}}{1 - a\,z^{-2}}$$

Note that this could have, equivalently, been written as:

$$P(z) = \frac{az^2}{1 - a\,z^2} + \frac{1}{1 - a\,z^{-2}}$$

□

Complex Spectral Density Functions

The inverse of the complex spectral and cross-spectral densities are given by the contour integral:

$$r_{xx}[\ell] = \frac{1}{2\pi j} \oint_C P_{xx}(z)\, z^{\ell-1}\, dz$$

$$r_{xy}[\ell] = \frac{1}{2\pi j} \oint_C P_{xy}(z)\, z^{\ell-1}\, dz$$

where the contour of integration C is to be taken counterclockwise and in the region of convergence.

Complex Spectral Density Functions

The inverse of the complex spectral and cross-spectral densities are given by the contour integral:

$$r_{xx}[\ell] = \frac{1}{2\pi j} \oint_C P_{xx}(z)\, z^{\ell-1}\, dz$$

$$r_{xy}[\ell] = \frac{1}{2\pi j} \oint_C P_{xy}(z)\, z^{\ell-1}\, dz$$

- In practice, these integrals are usually never performed, and tables, instead, are used.

Complex Spectral Density Functions

The inverse of the complex spectral and cross-spectral densities are given by the contour integral:

$$r_{xx}[\ell] = \frac{1}{2\pi j} \oint_C P_{xx}(z)\, z^{\ell-1}\, dz$$

$$r_{xy}[\ell] = \frac{1}{2\pi j} \oint_C P_{xy}(z)\, z^{\ell-1}\, dz$$

Some properties of the complex spectral densities include:

1. Conjugate-symmetry:

$$P_{xx}(z) = P_{xx}^*(1/z^*) \quad \text{and} \quad P_{xy}(z) = P_{yx}^*(1/z^*)$$

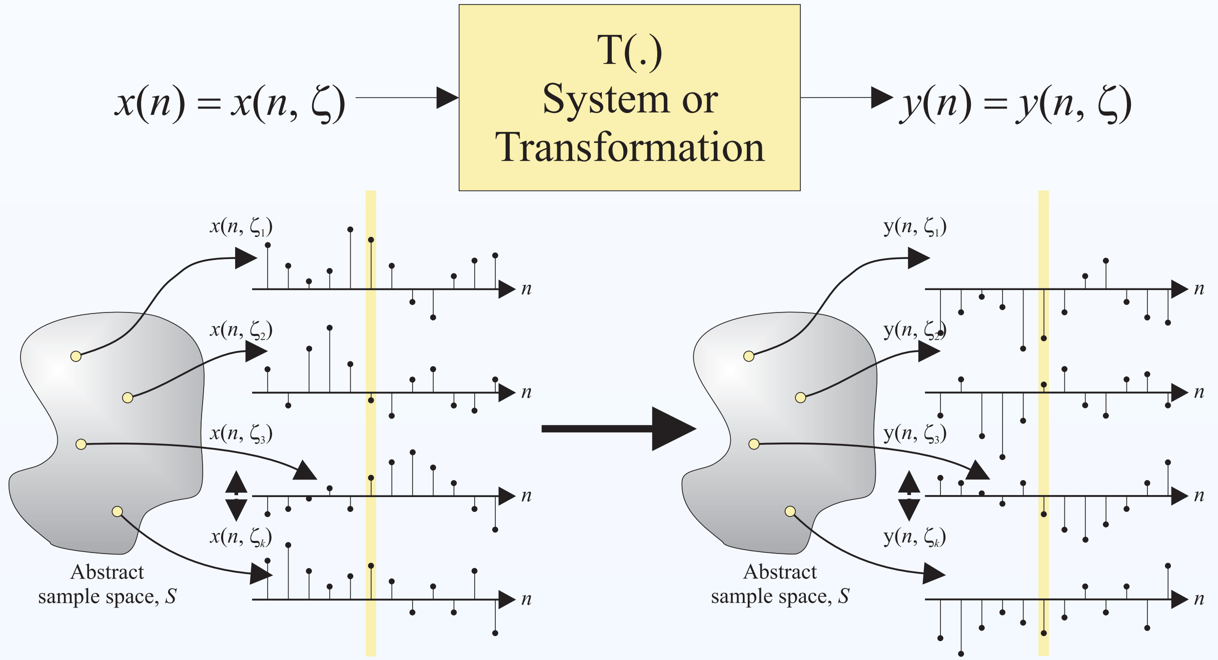
2. For the case when $x(n)$ is real, then:

$$P_{xx}(z) = P_{xx}(z^{-1})$$

Lecture Slideset 4
Linear Systems Theory

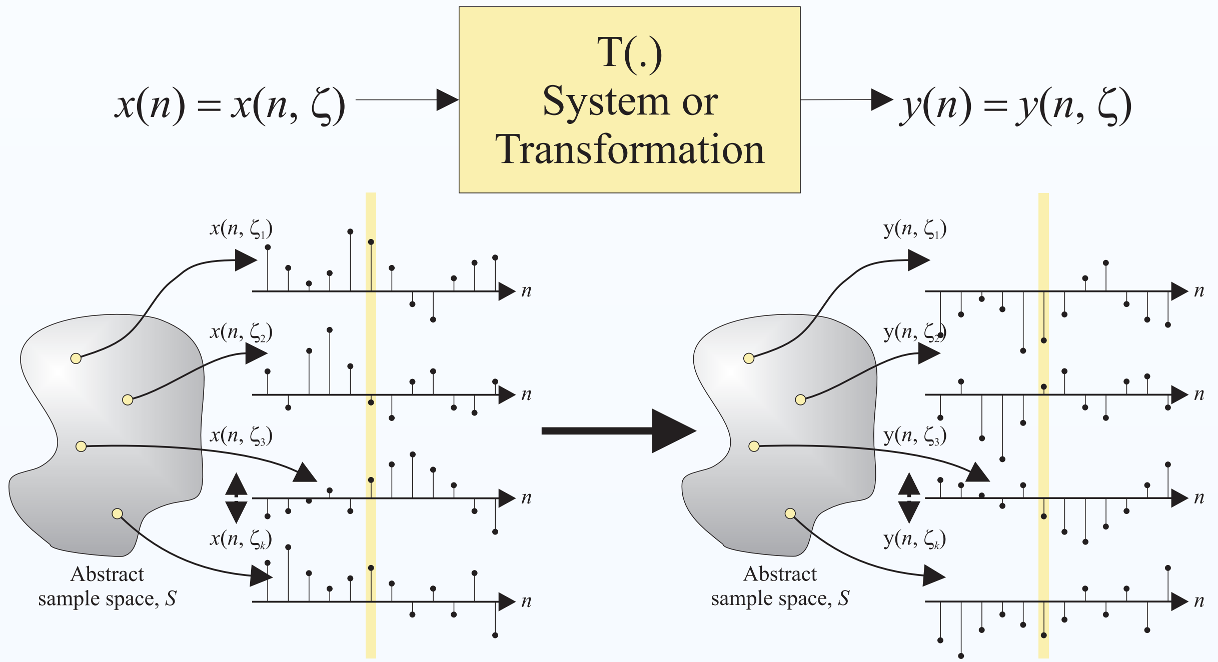
Systems with Stochastic Inputs

- Signal processing involves the transformation of signals to enhance certain characteristics; for example, to suppress noise, or to extract meaningful information.



A graphical representation of a random process at the output of a system in relation to a random process at the input of the system.

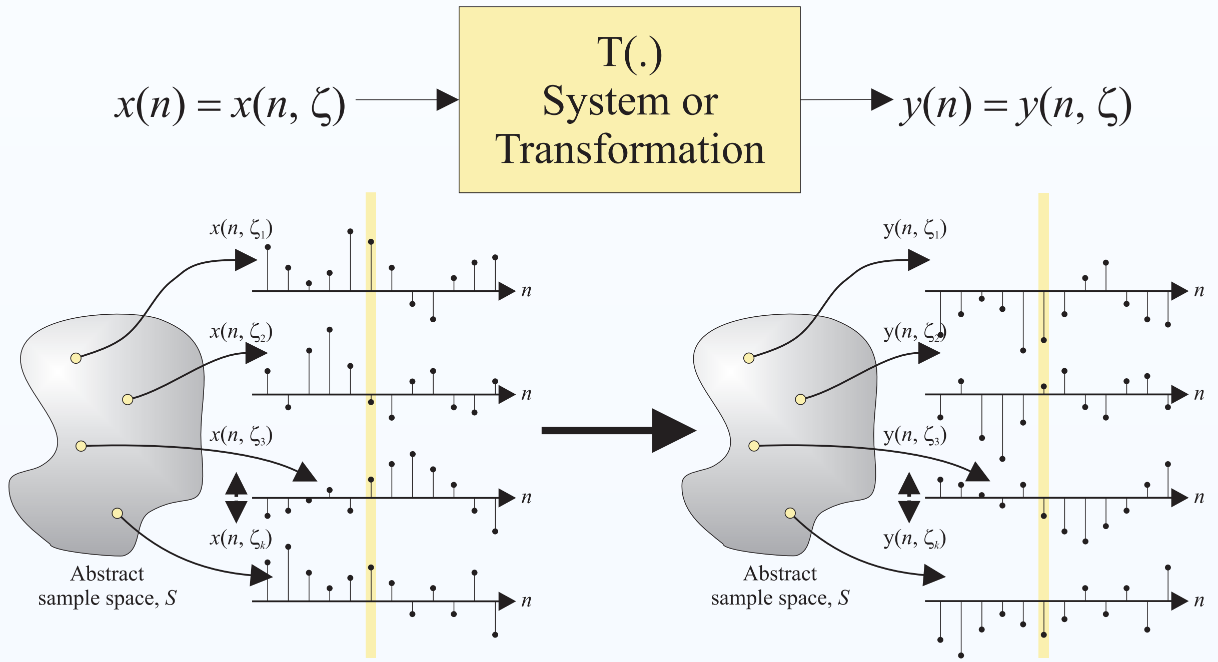
Systems with Stochastic Inputs



A graphical representation of a random process at the output of a system in relation to a random process at the input of the system.

- What does it mean to apply a stochastic signal to the input of a system?

Systems with Stochastic Inputs



A graphical representation of a random process at the output of a system in relation to a random process at the input of the system.

- What does it mean to apply a stochastic signal to the input of a system?
- This question is an interesting one since a stochastic process is not just a single sequence but an ensemble of sequences.

Systems with Stochastic Inputs

- In principle, the statistics of the output of any system can be expressed in terms of the statistics of the input.

Systems with Stochastic Inputs

- In principle, the statistics of the output of any system can be expressed in terms of the statistics of the input.
- However, in general this is a complicated problem except in special cases of particular types of signals or systems.

Systems with Stochastic Inputs

- In principle, the statistics of the output of any system can be expressed in terms of the statistics of the input.
- However, in general this is a complicated problem except in special cases of particular types of signals or systems.
- A special case is that of known-deterministic *linear systems*, and this is considered next.

Systems with Stochastic Inputs

- In principle, the statistics of the output of any system can be expressed in terms of the statistics of the input.
- However, in general this is a complicated problem except in special cases of particular types of signals or systems.
- A special case is that of known-deterministic *linear systems*, and this is considered next.
- In particular, if the input is a stationary stochastic process, and the system is linear time-invariant (LTI), then the statistics are even simpler.

Systems with Stochastic Inputs

- In principle, the statistics of the output of any system can be expressed in terms of the statistics of the input.
- However, in general this is a complicated problem except in special cases of particular types of signals or systems.
- A special case is that of known-deterministic *linear systems*, and this is considered next.
- In particular, if the input is a stationary stochastic process, and the system is linear time-invariant (LTI), then the statistics are even simpler.
- Moreover, it leads to a slightly simpler and intuitive explanation for the response of the system to the input.

Systems with Stochastic Inputs

- In principle, the statistics of the output of any system can be expressed in terms of the statistics of the input.
- However, in general this is a complicated problem except in special cases of particular types of signals or systems.
- A special case is that of known-deterministic *linear systems*, and this is considered next.
- In particular, if the input is a stationary stochastic process, and the system is linear time-invariant (LTI), then the statistics are even simpler.
- Moreover, it leads to a slightly simpler and intuitive explanation for the response of the system to the input.
- There are other systems that can be analysed, but due to time constraints, they are not considered in this course.

Systems with Stochastic Inputs

- In principle, the statistics of the output of any system can be expressed in terms of the statistics of the input.
- However, in general this is a complicated problem except in special cases of particular types of signals or systems.
- A special case is that of known-deterministic *linear systems*, and this is considered next.
- In particular, if the input is a stationary stochastic process, and the system is linear time-invariant (LTI), then the statistics are even simpler.
- Moreover, it leads to a slightly simpler and intuitive explanation for the response of the system to the input.
- There are other systems that can be analysed, but due to time constraints, they are not considered in this course.
- The case of random signals going through random systems is of great interest, but beyond the scope of this course.

Calculating Input-Output Statistics

	Time-domain	Frequency or transform domain
	LTI with stationary input	
Impulse response:	Manipulate convolution $y[n] = h[n] \star x[n] \Rightarrow$ $r_{yx}[\ell] = h[\ell] \star r_{xx}[\ell]$	Take z-transform of new convolution: $P_{yz}(z) = H(z)\, P_{xx}(z)$
Notes:	Solve convolution summation; Use graphical method.	Invert z-transform; Use partial fractions, tables,...
Difference equation:	Manipulate system difference equation: $\sum_{q=0}^{Q} a_p r_{yx}[\ell - q] = \sum_{p=0}^{P} b_p r_{xx}[\ell - p]$	Take z-transform of new equation: $P_{yx}(z) = P_{xx}(z) \frac{\sum_{p=0}^{P} b_p z^{-p}}{\sum_{q=0}^{Q} a_p z^{-q}}$
Notes:	Guess, e.g. $r_{yx}[\ell] = (\alpha\,\ell + \beta)\, r^{\ell}$ Recursive substitution.	Invert z-transform; Use partial fractions, tables, ...

Methods for solving the input-output statistics for a random signal passing through a deterministic linear system.

Calculating Input-Output Statistics

Example (Typical Question). A real-valued discrete-time random process $x[n]$ consists of independent and identically distributed (i. i. d.) random variables each with uniform density on the interval $[0, 6]$.

Calculating Input-Output Statistics

Example (Typical Question). A real-valued discrete-time random process $x[n]$ consists of independent and identically distributed (i. i. d.) random variables each with uniform density on the interval $[0, 6]$.

The process $x[n]$ is applied to a linear time-invariant (LTI) system with impulse response:

$$h[n] = \begin{cases} \left(\frac{2}{3}\right)^n, & n \geq 0 \\ 0, & n < 0 \end{cases}$$

$\bowtie$

Calculating Input-Output Statistics

Example (Typical Question). A real-valued discrete-time random process $x[n]$ consists of independent and identically distributed (i. i. d.) random variables each with uniform density on the interval $[0,\, 6]$.

The process $x[n]$ is applied to a linear time-invariant (LTI) system with impulse response:

$$h[n] = \begin{cases} \left(\frac{2}{3}\right)^n, & n \geq 0 \\ 0, & n < 0 \end{cases}$$

$\bowtie$

The output of this linear system is denoted as $y[n]$.

Calculating Input-Output Statistics

Example (Typical Question). A real-valued discrete-time random process $x[n]$ consists of independent and identically distributed (i. i. d.) random variables each with uniform density on the interval $[0, 6]$.

The process $x[n]$ is applied to a linear time-invariant (LTI) system with impulse response:

$$h[n] = \begin{cases} \left(\frac{2}{3}\right)^n, & n \geq 0 \\ 0, & n < 0 \end{cases} \qquad \bowtie$$

The output of this linear system is denoted as $y[n]$.

1. Calculate the output autocorrelation function $r_{yy}[\ell]$.

Calculating Input-Output Statistics

Example (Typical Question). A real-valued discrete-time random process $x[n]$ consists of independent and identically distributed (i. i. d.) random variables each with uniform density on the interval $[0, 6]$.

The process $x[n]$ is applied to a linear time-invariant (LTI) system with impulse response:

$$h[n] = \begin{cases} \left(\frac{2}{3}\right)^n, & n \geq 0 \\ 0, & n < 0 \end{cases}$$ ⋈

The output of this linear system is denoted as $y[n]$.

1. Calculate the output autocorrelation function $r_{yy}[\ell]$.

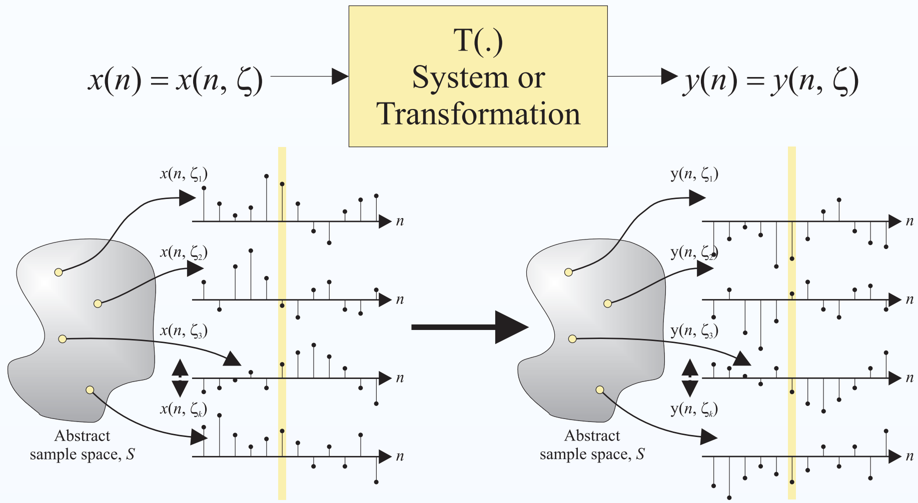
2. Suppose the i. i. d. process $x[n]$ now has a Weibull distribution with unit mean and variance of 3. Explain how your previous result might change, justifying your answer.

Calculating Input-Output Statistics

– End-of-Topic 56: **Summary of methods for calculating input-output statistics** –

Any Questions?

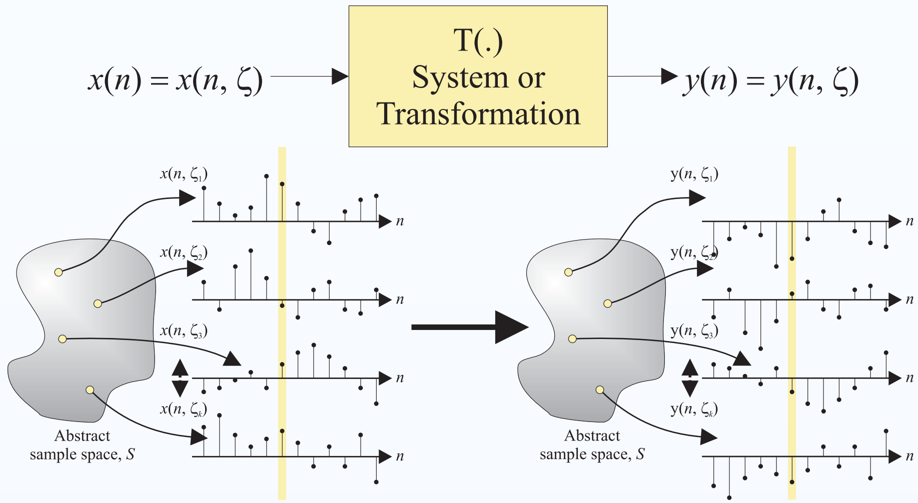
LTI Systems with Stationary Inputs



Since each sequence (realisation) of a stochastic process is a deterministic signal, there is a well-defined input signal producing a well-defined output signal corresponding to a single realisation of the output stochastic process:

$$y[n, \zeta] = \sum_{k=-\infty}^{\infty} h[k] \; x[n-k, \zeta]$$

LTI Systems with Stationary Inputs

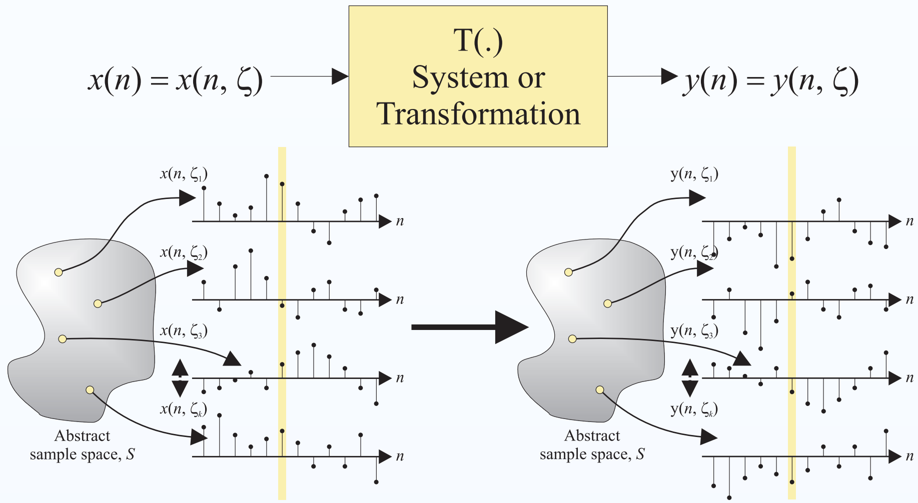


Since each sequence (realisation) of a stochastic process is a deterministic signal, there is a well-defined input signal producing a well-defined output signal corresponding to a single realisation of the output stochastic process:

$$y[n, \zeta] = \sum_{k=-\infty}^{\infty} h[k]\; x[n-k, \zeta]$$

- A complete description of $y[n, \zeta]$ requires the computation of an infinite number of convolutions, corresponding to each ζ.

LTI Systems with Stationary Inputs



Since each sequence (realisation) of a stochastic process is a deterministic signal, there is a well-defined input signal producing a well-defined output signal corresponding to a single realisation of the output stochastic process:

$$y[n, \zeta] = \sum_{k=-\infty}^{\infty} h[k]\, x[n-k, \zeta]$$

- Thus, better to consider the statistical properties of $y[n, \zeta]$ in terms of the statistical properties of the input and the system.

LTI Systems with Stationary Inputs

To investigate the statistical input-output properties of a linear system, note the following fundamental theorem:

Theorem (Expectation in Linear Systems). For any linear system,

$$\mathbb{E}\left[\mathcal{L}[x[n]]\right] = \mathcal{L}[\mathbb{E}\left[x[n]\right]]$$

$\diamondsuit$

LTI Systems with Stationary Inputs

To investigate the statistical input-output properties of a linear system, note the following fundamental theorem:

Theorem (Expectation in Linear Systems). For any linear system,

$$\mathbb{E}\left[\mathcal{L}[x[n]]\right] = \mathcal{L}[\mathbb{E}\left[x[n]\right]]$$

In other words, for example, the mean $\mu_y[n]$ of the output $y[n]$ equals the response of the system to the mean $\mu_x[n]$ of the input:

$$\mu_y[n] = \mathcal{L}[\mu_x[n]]$$

$\diamondsuit$

LTI Systems with Stationary Inputs

To investigate the statistical input-output properties of a linear system, note the following fundamental theorem:

Theorem (Expectation in Linear Systems). For any linear system,

$$\mathbb{E}\left[\mathcal{L}[x[n]]\right] = \mathcal{L}[\mathbb{E}\left[x[n]\right]]$$

In other words, for example, the mean $\mu_y[n]$ of the output $y[n]$ equals the response of the system to the mean $\mu_x[n]$ of the input:

$$\mu_y[n] = \mathcal{L}[\mu_x[n]]$$

- However, the definition extends to other statistics as well. $\diamondsuit$

LTI Systems with Stationary Inputs

To investigate the statistical input-output properties of a linear system, note the following fundamental theorem:

Theorem (Expectation in Linear Systems). For any linear system,

$$\mathbb{E}\left[\mathcal{L}[x[n]]\right] = \mathcal{L}[\mathbb{E}\left[x[n]\right]]$$

In other words, for example, the mean $\mu_y[n]$ of the output $y[n]$ equals the response of the system to the mean $\mu_x[n]$ of the input:

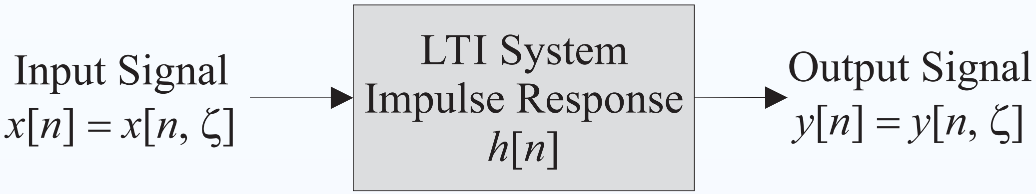
$$\mu_y[n] = \mathcal{L}[\mu_x[n]]$$

- However, the definition extends to other statistics as well. ♢

- Note, however, that while very useful, it is often more practical to derive most equations from first principals.

Input-output Statistics of a LTI System

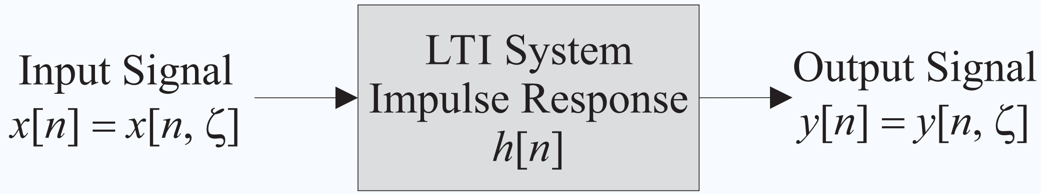
If a stationary stochastic process $x[n]$ with mean value μ_x and correlation $r_{xx}[\ell]$ is applied to the input of a LTI system with impulse response $h[n]$ and transfer function $H\left(e^{j\omega}\right)$, then the:



A linear time-invariant (LTI) system.

Input-output Statistics of a LTI System

If a stationary stochastic process $x[n]$ with mean value μ_x and correlation $r_{xx}[\ell]$ is applied to the input of a LTI system with impulse response $h[n]$ and transfer function $H\left(e^{j\omega}\right)$, then the:

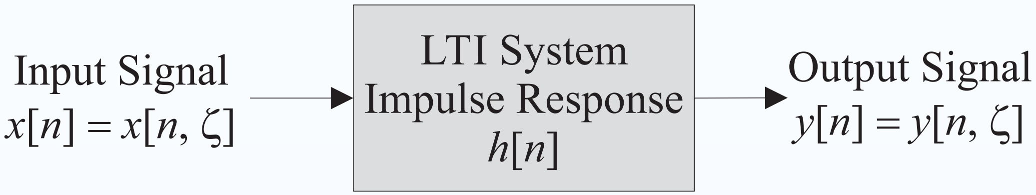


A linear time-invariant (LTI) system.

Output mean value is given by:

$$\mu_y = \mu_x \sum_{k=-\infty}^{\infty} h[k] = \mu_x H(e^{j0})$$

Input-output Statistics of a LTI System



Output mean value is given by:

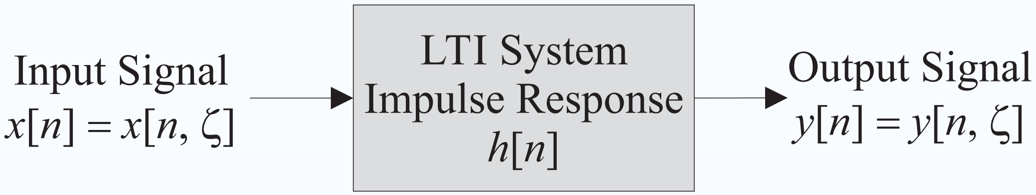
$$\mu_y = \mu_x \sum_{k=-\infty}^{\infty} h[k] = \mu_x \, H(e^{j0})$$

- This uses the linearity of the expectation operator:

$$\mu_y[n] = \mathbb{E}\left[\sum_{k=-\infty}^{\infty} h[k]\, x[n-k]\right] = \sum_{k=-\infty}^{\infty} h[k]\, \mathbb{E}\left[x[n-k]\right]$$

and since $x[n]$ is stationary, then $\mathbb{E}\left[x[n-k]\right] = \mu_x$.

Input-output Statistics of a LTI System



Output mean value is given by:

$$\mu_y = \mu_x \sum_{k=-\infty}^{\infty} h[k] = \mu_x H(e^{j0})$$

- This uses the linearity of the expectation operator:

$$\mu_y[n] = \mathbb{E}\left[\sum_{k=-\infty}^{\infty} h[k]\, x[n-k]\right] = \sum_{k=-\infty}^{\infty} h[k]\, \mathbb{E}\left[x[n-k]\right]$$

and since $x[n]$ is stationary, then $\mathbb{E}\left[x[n-k]\right] = \mu_x$. Since μ_x and $H(e^{j0})$ are constant, μ_y is also constant.

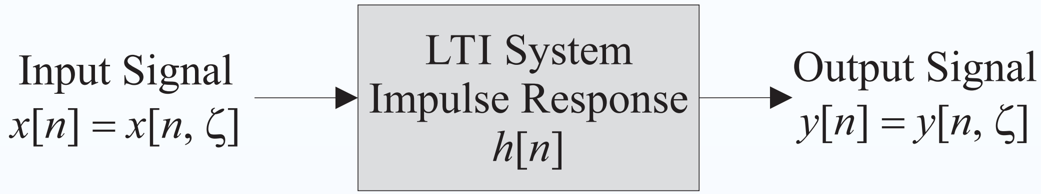
Input-output Statistics of a LTI System

Input Signal $x[n] = x[n, \zeta]$ → LTI System Impulse Response $h[n]$ → Output Signal $y[n] = y[n, \zeta]$

Input-output cross-correlation is given by:

$$r_{xy}[\ell] = h^*[-\ell] * r_{xx}[\ell] = \sum_{k=-\infty}^{\infty} h^*[-k] \, r_{xx}[\ell - k]$$

Input-output Statistics of a LTI System

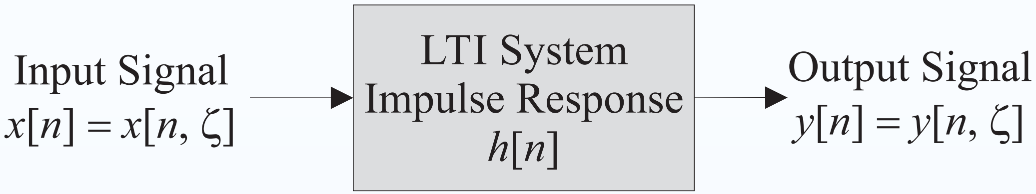


Input-output cross-correlation is given by:

$$r_{xy}[\ell] = h^*[-\ell] * r_{xx}[\ell] = \sum_{k=-\infty}^{\infty} h^*[-k]\, r_{xx}[\ell - k]$$

Similarly, $r_{yx}[\ell] = h[\ell] * r_{xx}[\ell]$, and is easy to prove:

Input-output Statistics of a LTI System



Input-output cross-correlation is given by:

$$r_{xy}[\ell] = h^*[-\ell] * r_{xx}[\ell] = \sum_{k=-\infty}^{\infty} h^*[-k] \, r_{xx}[\ell - k]$$

Similarly, $r_{yx}[\ell] = h[\ell] * r_{xx}[\ell]$, and is easy to prove:

$$r_{yx}[\ell] = \mathbb{E}\left[y[n] \; x^*[n - \ell] \right]$$

Input-output Statistics of a LTI System

Input Signal $x[n] = x[n, \zeta]$ → LTI System Impulse Response $h[n]$ → Output Signal $y[n] = y[n, \zeta]$

Input-output cross-correlation is given by:

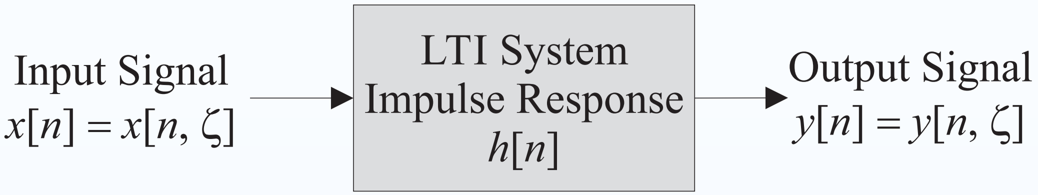
$$r_{xy}[\ell] = h^*[-\ell] * r_{xx}[\ell] = \sum_{k=-\infty}^{\infty} h^*[-k]\; r_{xx}[\ell - k]$$

Similarly, $r_{yx}[\ell] = h[\ell] * r_{xx}[\ell]$, and is easy to prove:

$$r_{yx}[\ell] = \mathbb{E}\left[y[n]\; x^*[n-\ell]\right]$$

$$= \mathbb{E}\left[\sum_{k=-\infty}^{\infty} h[k]\; x[n-k]\; x^*[n-\ell]\right]$$

Input-output Statistics of a LTI System



Input-output cross-correlation is given by:

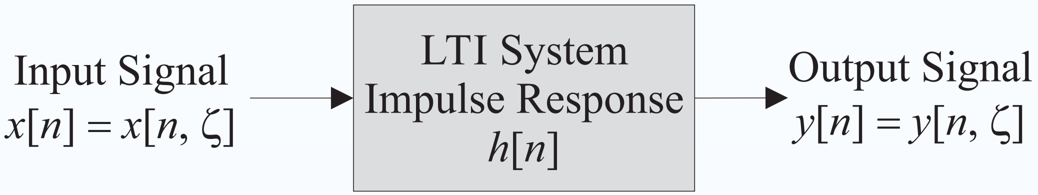
$$r_{xy}[\ell] = h^*[-\ell] * r_{xx}[\ell] = \sum_{k=-\infty}^{\infty} h^*[-k]\, r_{xx}[\ell - k]$$

Similarly, $r_{yx}[\ell] = h[\ell] * r_{xx}[\ell]$, and is easy to prove:

$$r_{yx}[\ell] = \mathbb{E}\left[y[n]\, x^*[n-\ell]\right]$$

$$= \sum_{k=-\infty}^{\infty} h[k]\, \mathbb{E}\left[x[n-k]\, x^*[n-\ell]\right]$$

Input-output Statistics of a LTI System



Input-output cross-correlation is given by:

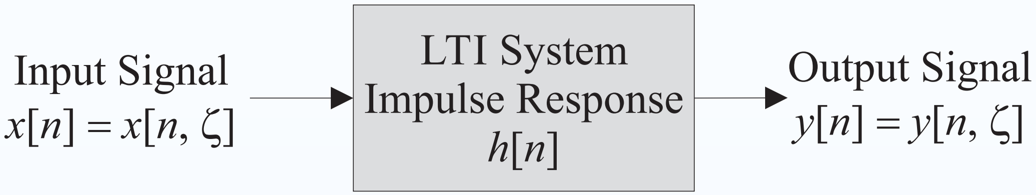
$$r_{xy}[\ell] = h^*[-\ell] * r_{xx}[\ell] = \sum_{k=-\infty}^{\infty} h^*[-k]\, r_{xx}[\ell - k]$$

Similarly, $r_{yx}[\ell] = h[\ell] * r_{xx}[\ell]$, and is easy to prove:

$$r_{yx}[\ell] = \mathbb{E}\left[y[n]\; x^*[n-\ell]\right]$$

$$= \sum_{k=-\infty}^{\infty} h[k]\, r_{xx}[\ell - k] = h[\ell] * r_{xx}[\ell]$$

Input-output Statistics of a LTI System

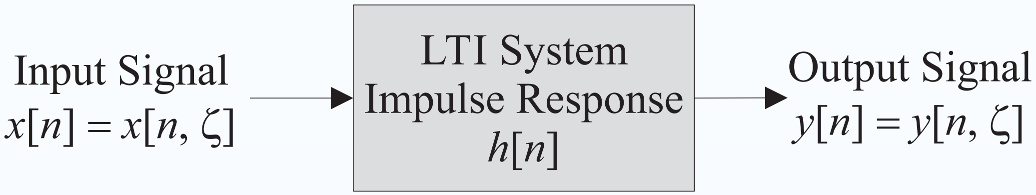


Output autocorrelation is obtained by post-multiplying the system-output by $y^*[n-\ell]$ and taking expectations:

$$r_{yy}[\ell] = \mathbb{E}\left[y[n]\; y^*[n-\ell]\right]$$

$$= \mathbb{E}\left[\sum_{k=-\infty}^{\infty} h[k]\; x[n-k]\; y^*[n-\ell]\right]$$

Input-output Statistics of a LTI System



Output autocorrelation is obtained by post-multiplying the system-output by $y^*[n-\ell]$ and taking expectations:

$$r_{yy}[\ell] = \mathbb{E}\left[y[n]\; y^*[n-\ell]\right]$$
$$= \mathbb{E}\left[\sum_{k=-\infty}^{\infty} h[k]\; x[n-k]\; y^*[n-\ell]\right]$$

and applying the linearity of the expectation operator:

$$r_{yy}[\ell] = \sum_{k=-\infty}^{\infty} h[k]\; \mathbb{E}\left[x[n-k]\; y^*[n-\ell]\right] = h[\ell] * r_{xy}[\ell]$$

Input-output Statistics of a LTI System

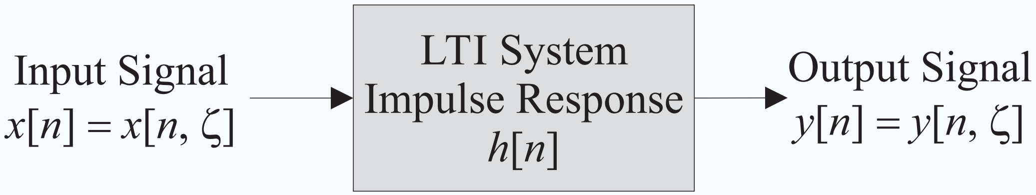
Input Signal $x[n] = x[n, \zeta]$ → LTI System Impulse Response $h[n]$ → Output Signal $y[n] = y[n, \zeta]$

Output autocorrelation is obtained by post-multiplying the system-output by $y^*[n-\ell]$ and taking expectations:

$$r_{yy}[\ell] = \sum_{k=-\infty}^{\infty} h[k]\, \mathbb{E}\left[x[n-k]\; y^*[n-\ell]\right] = h[\ell] * r_{xy}[\ell]$$

Substituting the expression for $r_{xy}[\ell] = h^*[-\ell] * r_{xx}[\ell]$ gives:

Input-output Statistics of a LTI System



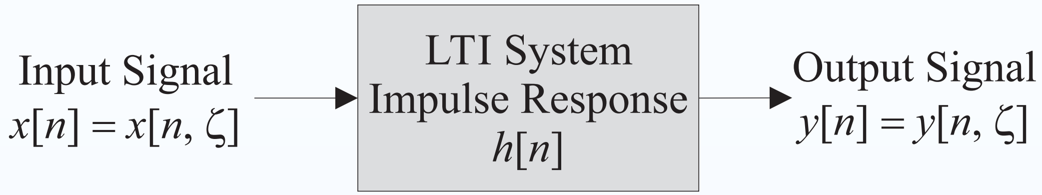
Output autocorrelation is obtained by post-multiplying the system-output by $y^*[n-\ell]$ and taking expectations:

$$r_{yy}[\ell] = \sum_{k=-\infty}^{\infty} h[k] \, \mathbb{E}\left[x[n-k] \; y^*[n-\ell]\right] = h[\ell] * r_{xy}[\ell]$$

Substituting the expression for $r_{xy}[\ell] = h^*[-\ell] * r_{xx}[\ell]$ gives:

$$r_{yy}[\ell] = h[\ell] * h^*[-\ell] * r_{xx}[\ell] = r_{hh}[\ell] * r_{xx}[\ell]$$

Input-output Statistics of a LTI System



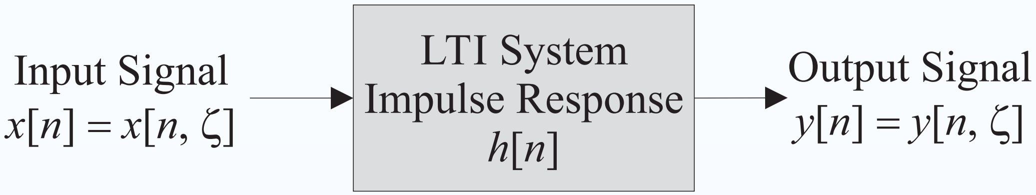
Output autocorrelation is obtained by post-multiplying the system-output by $y^*[n - \ell]$ and taking expectations:

$$r_{yy}[\ell] = h[\ell] * h^*[-\ell] * r_{xx}[\ell] = r_{hh}[\ell] * r_{xx}[\ell]$$

where $r_{hh}[\ell]$ is the *autocorrelation* of the impulse response:

$$r_{hh}[\ell] \triangleq h[\ell] * h^*[-\ell] = \sum_{n=-\infty}^{\infty} h[n]\; h^*[n - \ell]$$

Input-output Statistics of a LTI System

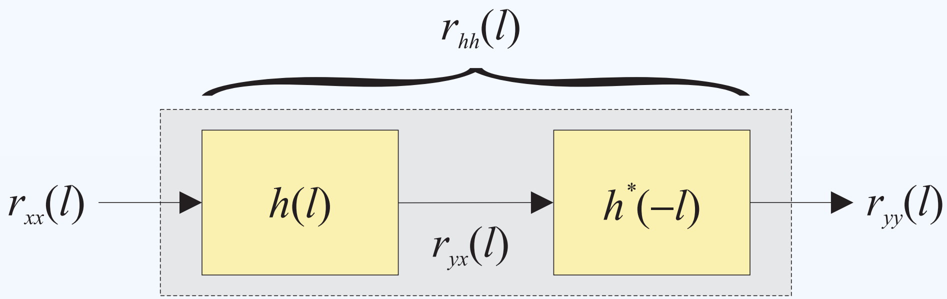


Output autocorrelation is given by the expression:

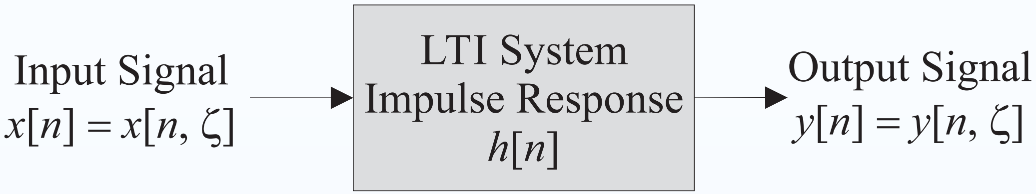
$$r_{yy}[\ell] = h[\ell] * h^*[-\ell] * r_{xx}[\ell] = r_{hh}[\ell] * r_{xx}[\ell]$$

where $r_{hh}[\ell]$ is the *autocorrelation* of the impulse response:

$$r_{hh}[\ell] \triangleq h[\ell] * h^*[-\ell] = \sum_{n=-\infty}^{\infty} h[n]\ h^*[n-\ell]$$



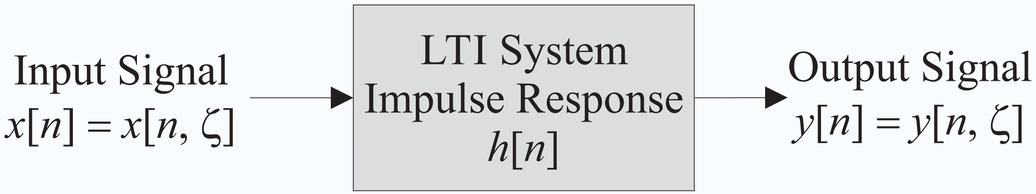
Input-output Statistics of a LTI System



Output-power of the process $y[n]$ is given by $r_{yy}[0] = \mathbb{E}\left[|y[n]|^2\right]$,

and therefore since $r_{yy}[\ell] = r_{hh}[\ell] * r_{xx}[\ell]$,

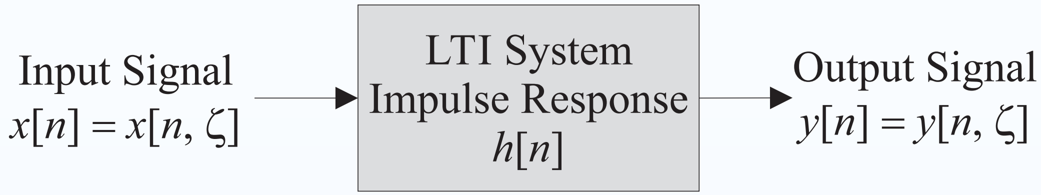
Input-output Statistics of a LTI System



Output-power of the process $y[n]$ is given by $r_{yy}[0] = \mathbb{E}\left[|y[n]|^2\right]$, and therefore since $r_{yy}[\ell] = r_{hh}[\ell] * r_{xx}[\ell]$,

$$P_{yy} = r_{hh}[\ell] * r_{xx}[\ell]|_{\ell=0} = \sum_{k=-\infty}^{\infty} r_{hh}[k]\; r_{xx}[-k]$$

Input-output Statistics of a LTI System



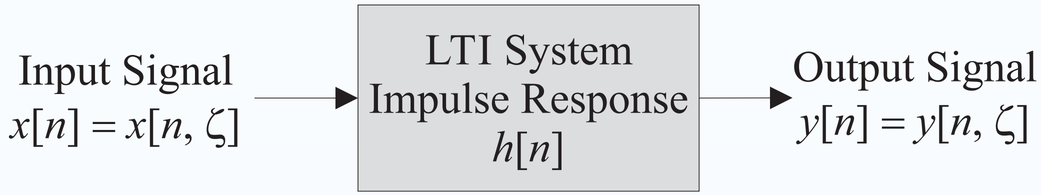
Output-power of the process $y[n]$ is given by $r_{yy}[0] = \mathbb{E}\left[|y[n]|^2\right]$, and therefore since $r_{yy}[\ell] = r_{hh}[\ell] * r_{xx}[\ell]$,

$$P_{yy} = r_{hh}[\ell] * r_{xx}[\ell]|_{\ell=0} = \sum_{k=-\infty}^{\infty} r_{hh}[k]\, r_{xx}[-k]$$

Noting power, P_{yy}, is real, then taking complex-conjugates using $r_{xx}^*[-\ell] = r_{xx}[\ell]$:

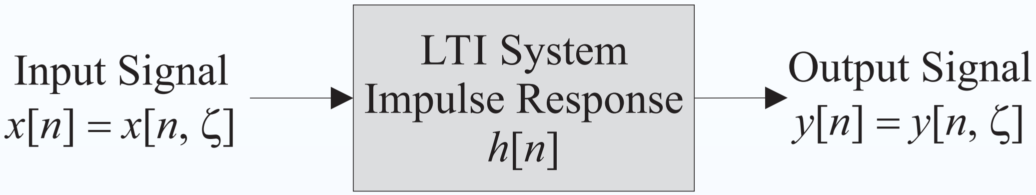
$$P_{yy} = \sum_{k=-\infty}^{\infty} r_{hh}^*[k]\, r_{xx}[k] = \sum_{n=-\infty}^{\infty} h^*[n] \sum_{k=-\infty}^{\infty} r_{xx}[n+k]\, h[k]$$

Input-output Statistics of a LTI System



Output pdf It, in general, it is very difficult to calculate the pdf of the output of a LTI system, except in special cases, namely Gaussian processes.

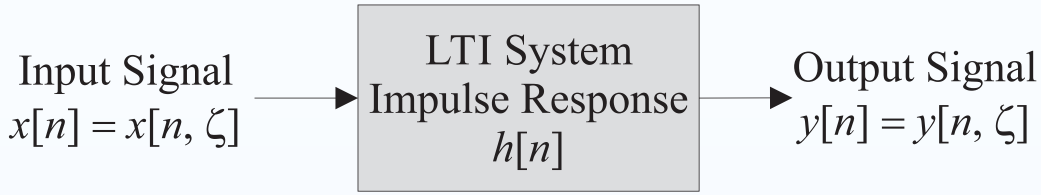
Input-output Statistics of a LTI System



Output pdf It is difficult to calculate the pdf of the output of a LTI system, except in special cases, namely Gaussian processes.

- Finally, note that the covariance sequences is just the correlation sequences with the mean removed.
- As a result, the covariance functions satisfy a set of equations analogous to those derived above.

Input-output Statistics of a LTI System

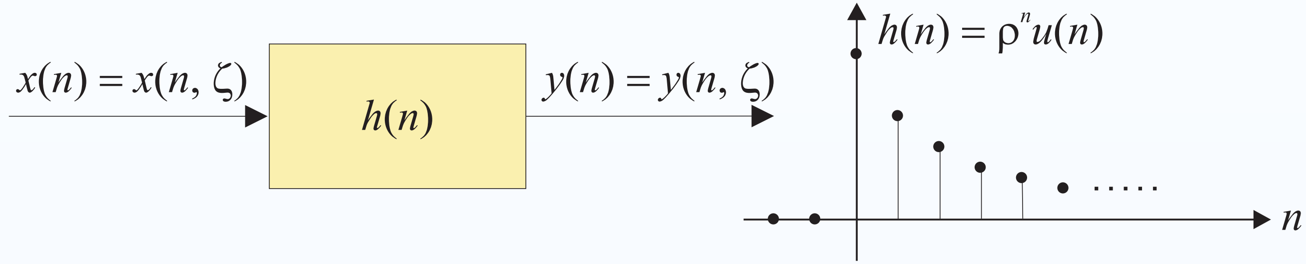


Output pdf It is difficult to calculate the pdf of the output of a LTI system, except in special cases, namely Gaussian processes.

- Finally, note that the covariance sequences is just the correlation sequences with the mean removed.
- As a result, the covariance functions satisfy these equations:

$$
\begin{aligned}
\gamma_{yx}[\ell] &= h[\ell] * \gamma_{xx}[\ell] \\
\gamma_{xy}[\ell] &= h^*[-\ell] * \gamma_{xx}[\ell] \\
\gamma_{yy}[\ell] &= h[\ell] * \gamma_{xy}[\ell] \\
&= h[\ell] * h^*[-\ell] * \gamma_{xx}[\ell]
\end{aligned}
$$

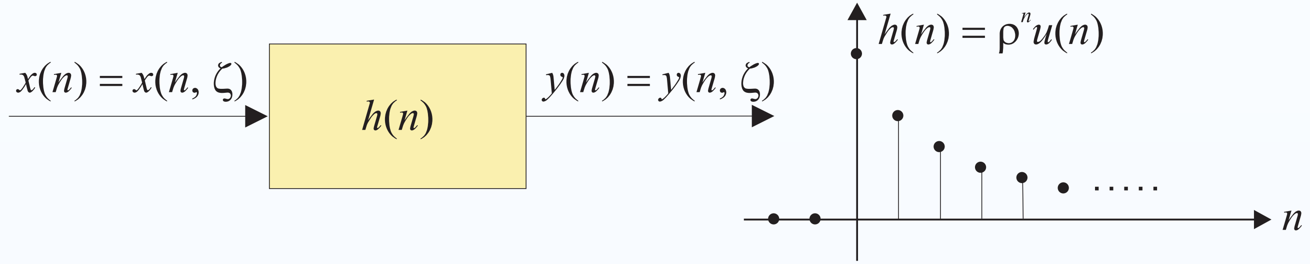
Input-output Statistics of a LTI System



Example (Simple example). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2 \delta[\ell]$

- Calculate the mean, autocorrelation and autocovariance sequences of the output, $y[n]$, as well as the cross-correlation and cross-covariance functions between the input and the output. ⋈

Input-output Statistics of a LTI System



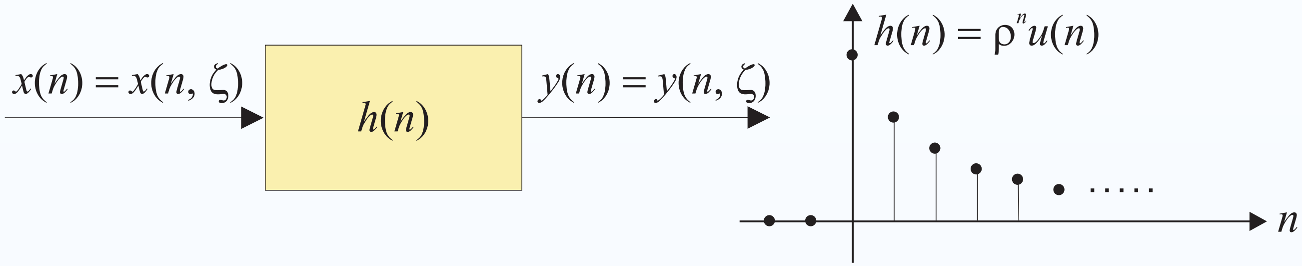
Example (Simple example). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2 \delta[\ell]$

- Calculate the mean, autocorrelation and autocovariance sequences of the output, $y[n]$, as well as the cross-correlation and cross-covariance functions between the input and the output.

SOLUTION. **Output mean value** First, calculate the mean.

$$\mu_y = \mu_x \sum_{k=-\infty}^{\infty} h[k] = \mu_x \sum_{k=0}^{\infty} \rho^k = \frac{\mu_x}{1-\rho} \qquad \square$$

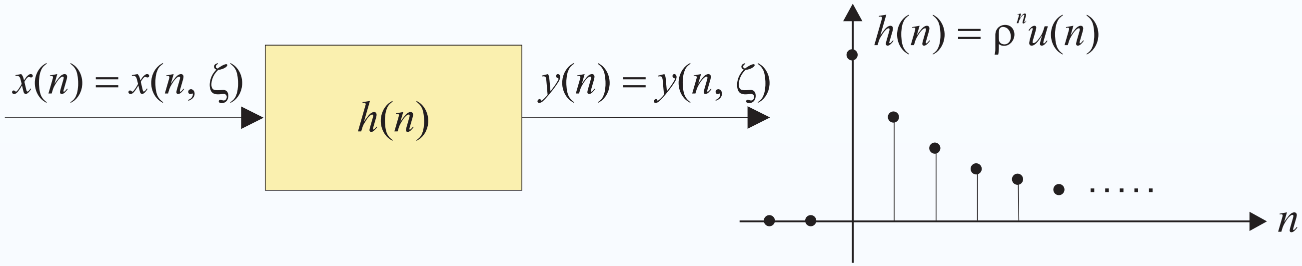
Input-output Statistics of a LTI System



Example (Simple example). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2 \delta[\ell]$

SOLUTION. **Input-output cross-covariance** Since the input and the output both have nonzero mean, then it is easiest to first calculate the auto- and cross-*covariance* functions, and then use these to find the auto- and cross-*correlation* functions.

Input-output Statistics of a LTI System



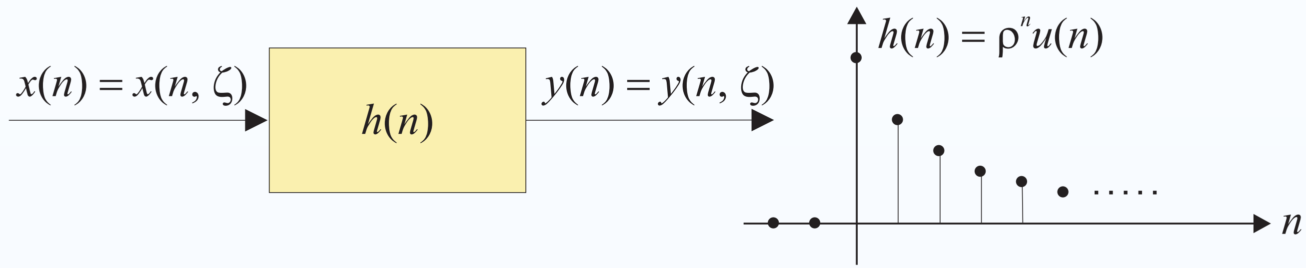
Example (Simple example). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2 \delta[\ell]$

SOLUTION. **Input-output cross-covariance** Thus, the output-input cross-covariance is given by:

$$\gamma_{yx}[\ell] = h[\ell] * \gamma_{xx}[\ell] = \left(\rho^\ell u[\ell]\right) * \left(\sigma_x^2 \delta[\ell]\right) = \sigma_x^2 \rho^\ell u[\ell]$$

□

Input-output Statistics of a LTI System



Example (Simple example). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2 \delta[\ell]$

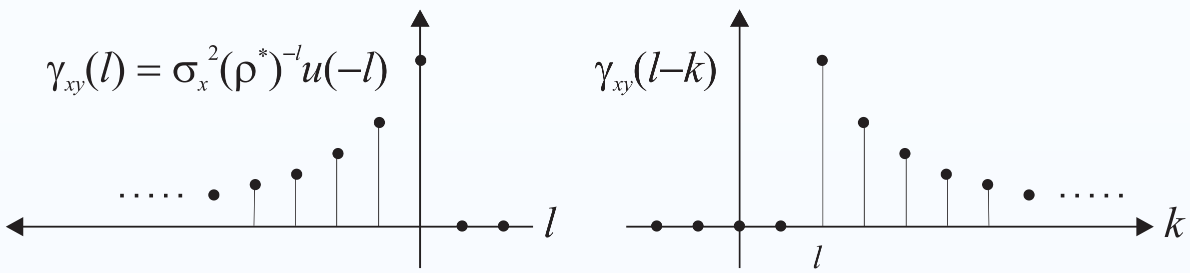
SOLUTION. **Input-output cross-covariance** Thus, the output-input cross-covariance is given by:

$$\gamma_{yx}[\ell] = h[\ell] * \gamma_{xx}[\ell] = \left(\rho^\ell u[\ell]\right) * \left(\sigma_x^2 \delta[\ell]\right) = \sigma_x^2 \rho^\ell u[\ell]$$

The input-output cross-covariance is

$$\gamma_{xy}[\ell] = \gamma_{yx}^*[-\ell] = \sigma_x^2 (\rho^*)^{-\ell} u[-\ell] \qquad \square$$

Input-output Statistics of a LTI System



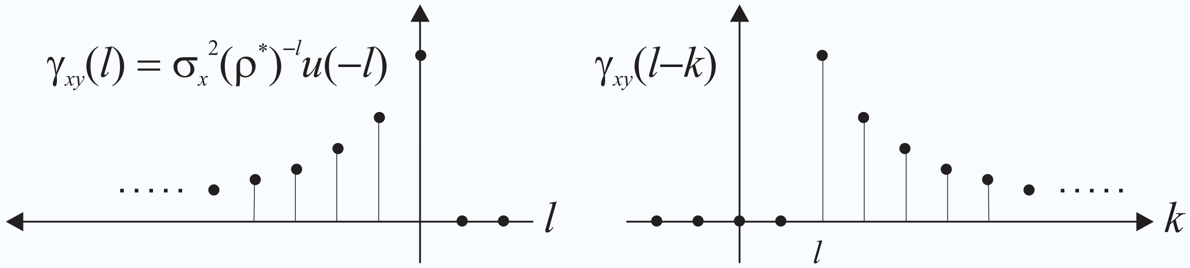
Example (Simple example). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2 \delta[\ell]$

SOLUTION. **Output autocovariance** Next:

$$\gamma_{yy}[\ell] = h[\ell] * \gamma_{xy}[\ell] = \sum_{k=-\infty}^{\infty} h[k]\, \gamma_{xy}[\ell - k]$$

□

Input-output Statistics of a LTI System



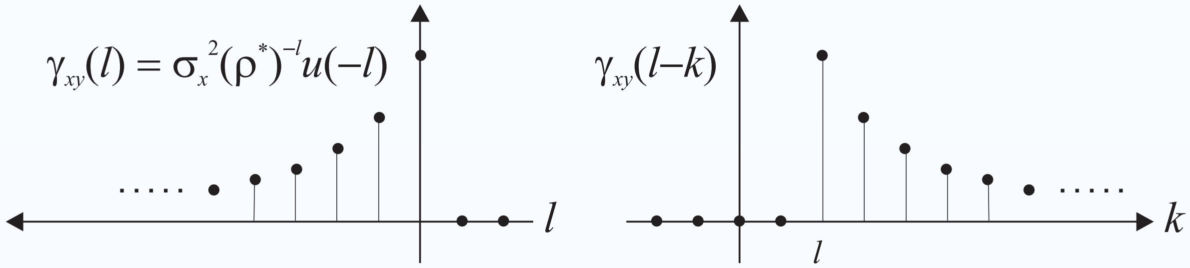
Example (Simple example). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2 \delta[\ell]$

SOLUTION. **Output autocovariance** Next:

$$\gamma_{yy}[\ell] = h[\ell] * \gamma_{xy}[\ell] = \sum_{k=-\infty}^{\infty} h[k]\, \gamma_{xy}[\ell - k] \qquad \square$$

The input-output cross-covariance sequence, $\gamma_{xy}[\ell]$, is plotted, along with $\gamma_{xy}[\ell - k]$ as a function of k.

Input-output Statistics of a LTI System



Example (Simple example). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2 \delta[\ell]$

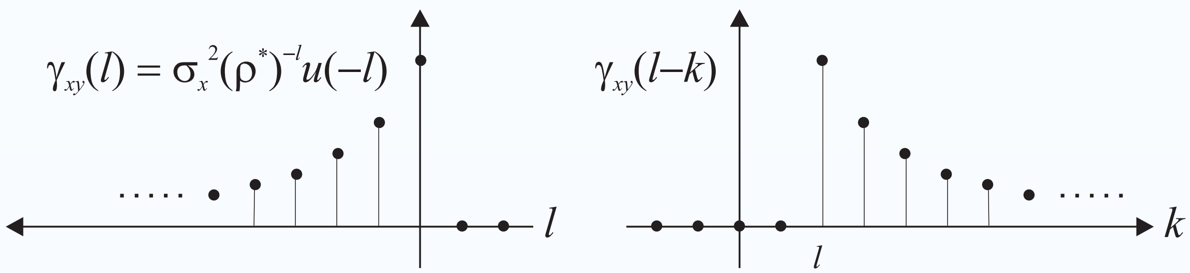
SOLUTION. **Output autocovariance** Next:

$$\gamma_{yy}[\ell] = h[\ell] * \gamma_{xy}[\ell] = \sum_{k=-\infty}^{\infty} h[k] \; \gamma_{xy}[\ell - k]$$

Hence, if $\ell > 0$ it follows

$$\gamma_{yy}[\ell] = \sum_{k=\ell}^{\infty} \rho^k \, \sigma_x^2 (\rho^*)^{-(\ell-k)}$$

Input-output Statistics of a LTI System



Example (Simple example). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2 \delta[\ell]$

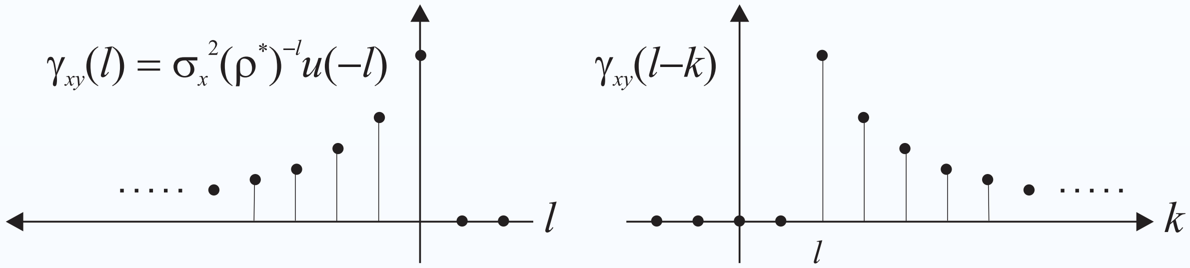
SOLUTION. **Output autocovariance** Hence, if $\ell > 0$ it follows

$$\gamma_{yy}[\ell] = \sum_{k=\ell}^{\infty} \rho^k \, \sigma_x^2 (\rho^*)^{-(\ell-k)}$$

Substituting $m = k - \ell$, so $k = \{\ell, \infty\} \Rightarrow m = \{0, \infty\}$:

$$\gamma_{yy}[\ell] = \sigma_x^2 \sum_{m=0}^{\infty} \rho^\ell \, \rho^m \, (\rho^*)^m$$

Input-output Statistics of a LTI System



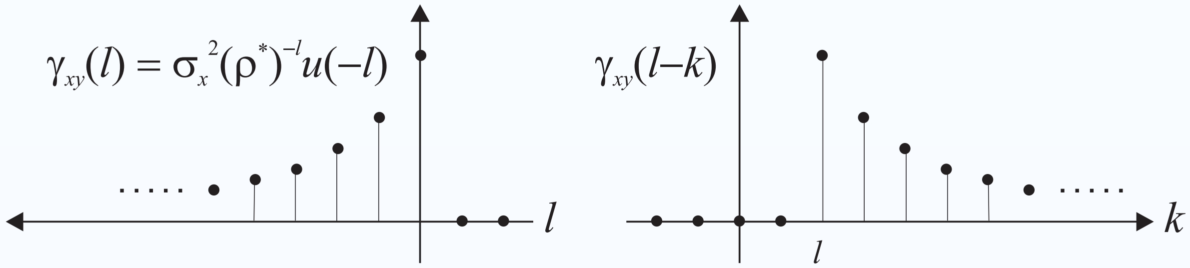
Example (Simple example). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2 \delta[\ell]$

SOLUTION. **Output autocovariance** Substituting $m = k - \ell$, so $k = \{\ell, \infty\} \Rightarrow m = \{0, \infty\}$:

$$\gamma_{yy}[\ell] = \sigma_x^2 \sum_{m=0}^{\infty} \rho^\ell \, \rho^m \, (\rho^*)^m$$

$$= \sigma_x^2 \rho^\ell \sum_{m=0}^{\infty} \left(|\rho|^2\right)^m = \frac{\sigma_x^2 \rho^\ell}{1 - |\rho|^2}, \; \ell > 0$$

Input-output Statistics of a LTI System



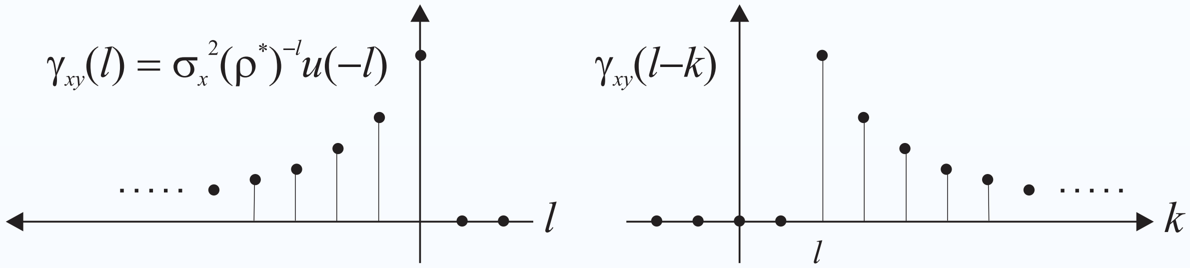
Example (Simple example). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2 \delta[\ell]$

SOLUTION. **Output autocovariance** If $\ell \leq 0$, then the summation is slightly different:

$$\gamma_{yy}[\ell] = \sum_{k=0}^{\infty} \rho^k \, \sigma_x^2 (\rho^*)^{-(\ell-k)}$$

$$= \sigma_x^2 (\rho^*)^{-\ell} \sum_{k=0}^{\infty} \left(|\rho|^2\right)^k = \frac{\sigma_x^2 (\rho^*)^{-\ell}}{1-|\rho|^2}, \; \ell \leq 0 \qquad \square$$

Input-output Statistics of a LTI System



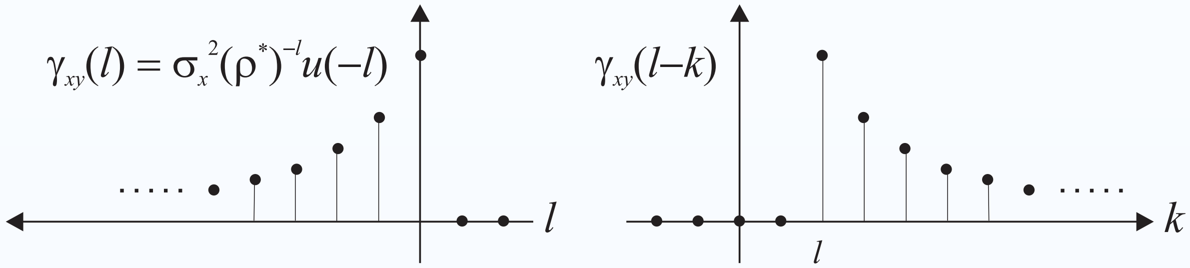
Example (Simple example). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2 \delta[\ell]$

SOLUTION. **Input-output cross-correlation** This can now be calculated using the relationship:

$$r_{xy}[\ell] = \gamma_{xy}[\ell] + \mu_x \, \mu_y^*$$

□

Input-output Statistics of a LTI System



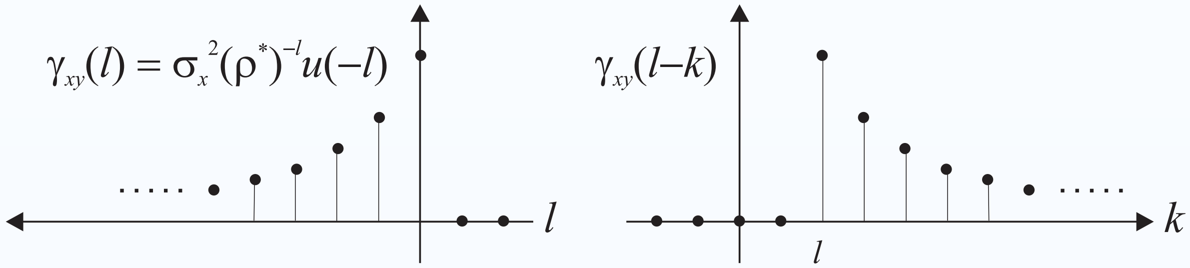
Example (Simple example). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2 \delta[\ell]$

SOLUTION. **Input-output cross-correlation** This can now be calculated using the relationship:

$$
\begin{aligned}
r_{xy}[\ell] &= \gamma_{xy}[\ell] + \mu_x \, \mu_y^* \\
&= \sigma_x^2 (\rho^*)^{-\ell} u[-\ell] + \mu_x \, \frac{\mu_x^*}{1 - \rho^*}
\end{aligned}
$$

□

Input-output Statistics of a LTI System



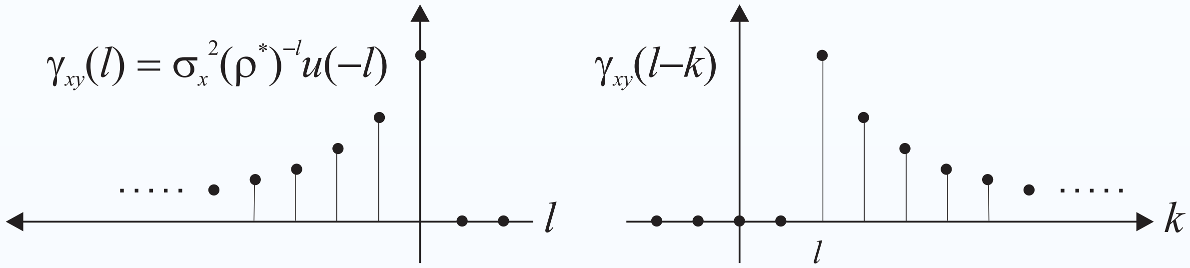
Example (Simple example). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2 \delta[\ell]$

SOLUTION. **Input-output cross-correlation** This can now be calculated using the relationship:

$$
\begin{aligned}
r_{xy}[\ell] &= \gamma_{xy}[\ell] + \mu_x \, \mu_y^* \\
&= \sigma_x^2 (\rho^*)^{-\ell} u[-\ell] + \mu_x \frac{\mu_x^*}{1-\rho^*} \\
&= \sigma_x^2 (\rho^*)^{-\ell} u[-\ell] + \frac{|\mu_x|^2}{1-\rho^*}
\end{aligned}
$$

□

Input-output Statistics of a LTI System

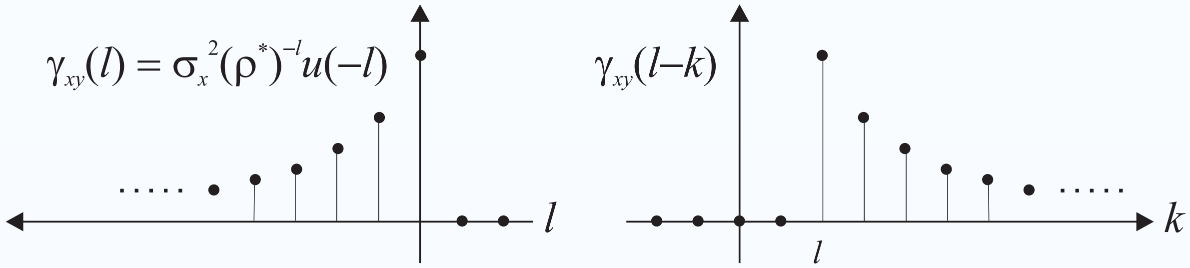


Example (Simple example). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2 \delta[\ell]$

SOLUTION. **Output autocorrelation** In a similar manner, the autocorrelation of the output is given by:

$$r_{yy}[\ell] = \gamma_{yy}[\ell] + |\mu_y|^2 = \begin{cases} \dfrac{\sigma_x^2 \rho^\ell}{1-|\rho|^2} + \left|\dfrac{\mu_x}{1-\rho}\right|^2 & \ell > 0 \\ \dfrac{\sigma_x^2 (\rho^*)^{-\ell}}{1-|\rho|^2} + \left|\dfrac{\mu_x}{1-\rho}\right|^2 & \ell \leq 0 \end{cases} \qquad \square$$

Input-output Statistics of a LTI System



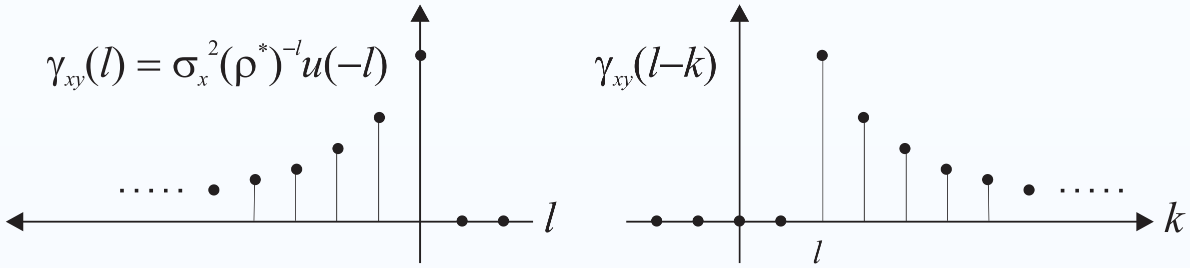
Example (Simple example). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2 \delta[\ell]$

SOLUTION. **Output autocorrelation** In a similar manner:

$$r_{yy}[\ell] = \gamma_{yy}[\ell] + |\mu_y|^2 = \begin{cases} \frac{\sigma_x^2 \rho^\ell}{1-|\rho|^2} + \left|\frac{\mu_x}{1-\rho}\right|^2 & \ell > 0 \\ \frac{\sigma_x^2 (\rho^*)^{-\ell}}{1-|\rho|^2} + \left|\frac{\mu_x}{1-\rho}\right|^2 & \ell \leq 0 \end{cases}$$

- Note that these results show that a process with the exponential correlation function can always be generated by applying white noise to a stable first-order system. □

Input-output Statistics of a LTI System



Example (Simple example). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2 \delta[\ell]$

SOLUTION. **Output autocorrelation** In a similar manner:

$$r_{yy}[\ell] = \gamma_{yy}[\ell] + |\mu_y|^2 = \begin{cases} \dfrac{\sigma_x^2 \rho^\ell}{1-|\rho|^2} + \left|\dfrac{\mu_x}{1-\rho}\right|^2 & \ell > 0 \\ \dfrac{\sigma_x^2 (\rho^*)^{-\ell}}{1-|\rho|^2} + \left|\dfrac{\mu_x}{1-\rho}\right|^2 & \ell \leq 0 \end{cases}$$

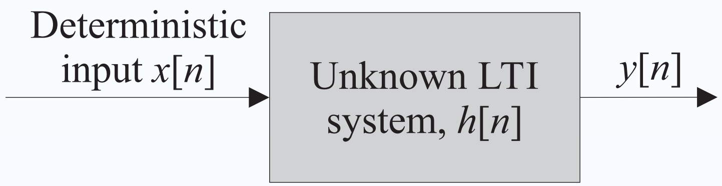
- More generally, it will be seen that wide-sense stationary of arbitrary autocorrelation sequence can be obtained by driving a LTI system by WGN. □

Input-output Statistics of a LTI System

– End-of-Topic 57: **Calculating input-output statistics in the time-domain with the system impulse response** –

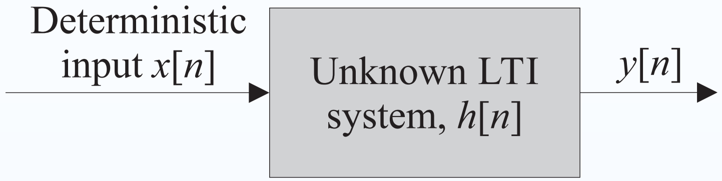
Any Questions?

System identification



What signals might be used for System Identification?

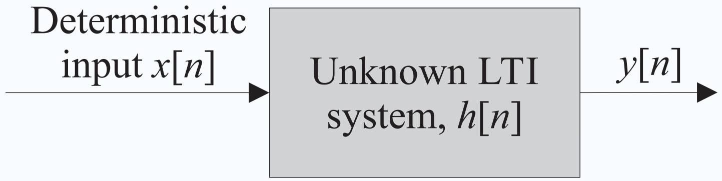
System identification



What signals might be used for System Identification?

There are three key methods from our deterministic signal analysis for system identification:

System identification

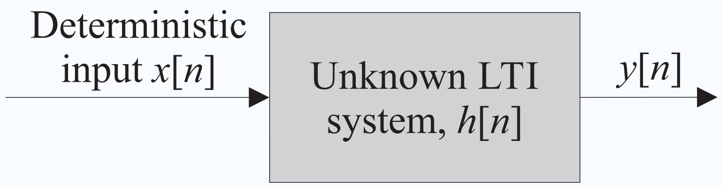


What signals might be used for System Identification?

There are three key methods from our deterministic signal analysis for system identification:

Impulse A simple input, but **difficult to generate**. The output is $y[n] = h[n]$, the system impulse response.

System identification



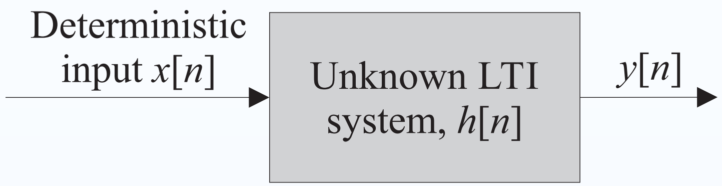
What signals might be used for System Identification?

There are three key methods from our deterministic signal analysis for system identification:

Impulse A simple input, but **difficult to generate**. The output is $y[n] = h[n]$, the system impulse response.

Step input A simple to generate signal, with the output $y[n] = \sum_{k=0}^{n} h[k]$ being the step response. The impulse response is obtained by taking the difference at the output.

System identification



What signals might be used for System Identification?

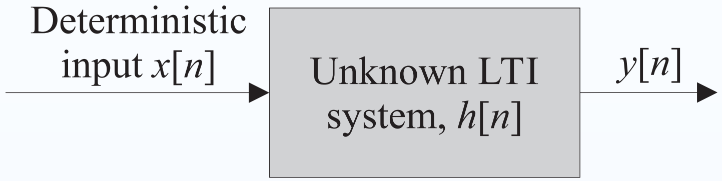
There are three key methods from our deterministic signal analysis for system identification:

Impulse A simple input, but **difficult to generate**. The output is $y[n] = h[n]$, the system impulse response.

Step input A simple to generate signal, with the output $y[n] = \sum_{k=0}^{n} h[k]$ being the step response. The impulse response is obtained by taking the difference at the output.

- This is problematic, as the difference signal can lead to errors when there is a small amount of noise in the signals.

System identification



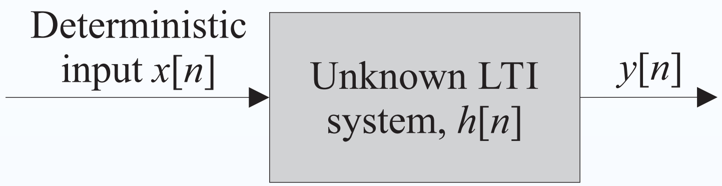
What signals might be used for System Identification?

There are three key methods from our deterministic signal analysis for system identification:

Harmonic input A simple to generate signal, $x[n] = \cos \omega_0 n$, leading to the output:

$$y[n] = \left| H\left(e^{j\omega_0}\right) \right| \cos\left(\omega_0 n + \arg H\left(e^{j\omega_0}\right)\right)$$

System identification



What signals might be used for System Identification?

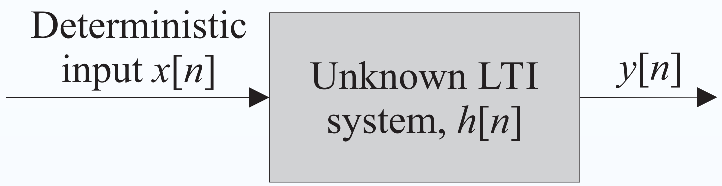
There are three key methods from our deterministic signal analysis for system identification:

Harmonic input A simple to generate signal, $x[n] = \cos \omega_0 n$, leading to the output:

$$y[n] = \left|H\left(e^{j\omega_0}\right)\right| \cos\left(\omega_0 n + \arg H\left(e^{j\omega_0}\right)\right)$$

- By sweeping across frequencies, the magnitude and phase response of $H\left(e^{j\omega}\right)$ can be calculated.

System identification



What signals might be used for System Identification?

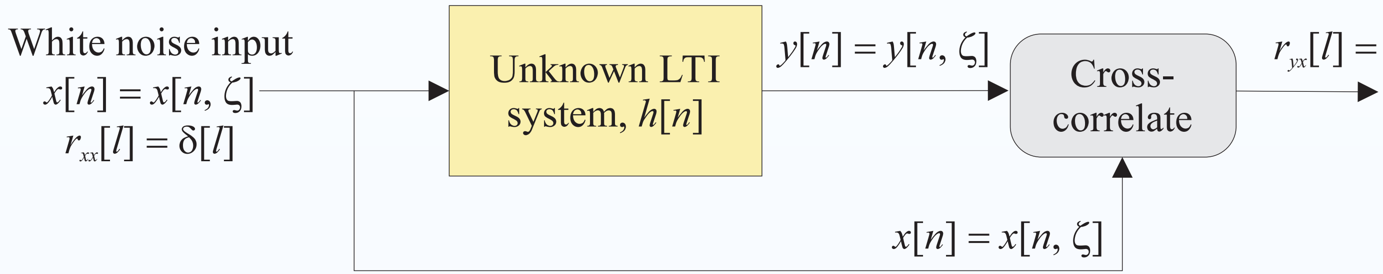
There are three key methods from our deterministic signal analysis for system identification:

Harmonic input A simple to generate signal, $x[n] = \cos \omega_0 n$, leading to the output:

$$y[n] = \left|H\left(e^{j\omega_0}\right)\right| \cos\left(\omega_0 n + \arg H\left(e^{j\omega_0}\right)\right)$$

- By sweeping across frequencies, the magnitude and phase response of $H\left(e^{j\omega}\right)$ can be calculated.
- The inverse-DTFT can then be used to reconstruct the impulse response, $h[n]$.

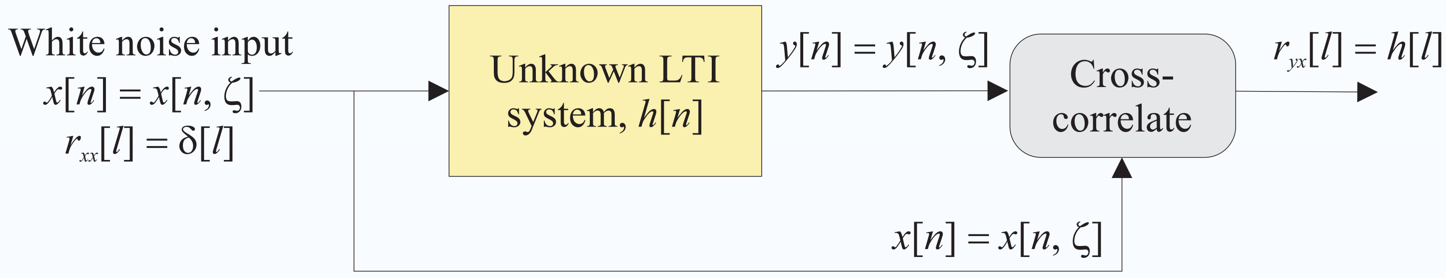
System identification



System identification by cross-correlation.

The input-output cross-correlation of a LTI system is the basis for a classical method of identification of an unknown linear system.

System identification



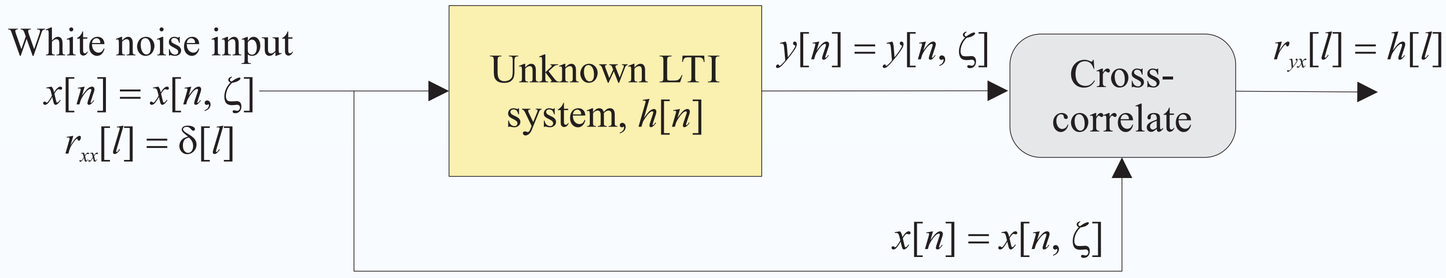
System identification by cross-correlation.

The input-output cross-correlation of a LTI system is the basis for a classical method of identification of an unknown linear system.

The system is excited with a WGN input with ACS:

$$r_{xx}[\ell] = \delta[\ell]$$

System identification



The input-output cross-correlation of a LTI system is the basis for a classical method of identification of an unknown linear system.

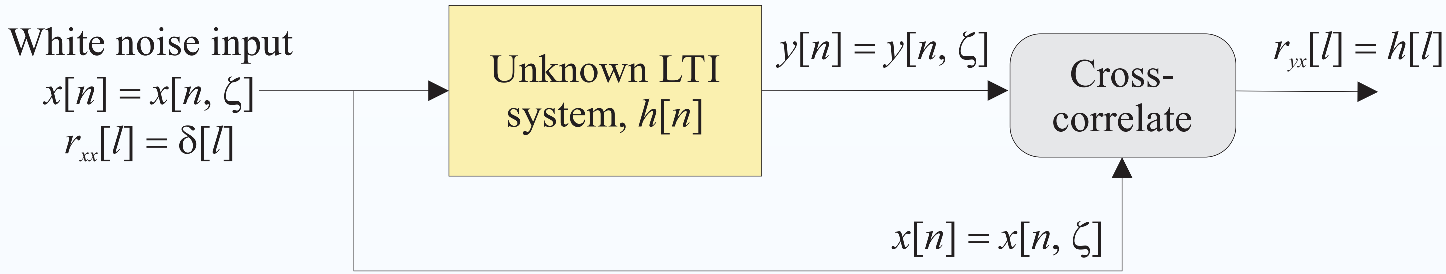
The system is excited with a WGN input with ACS:

$$r_{xx}[\ell] = \delta[\ell]$$

Since the output-input cross-correlation can be written as:

$$r_{yx}[\ell] = h[\ell] * r_{xx}[\ell]$$

System identification



The input-output cross-correlation of a LTI system is the basis for a classical method of identification of an unknown linear system.

The system is excited with a WGN input with ACS:

$$r_{xx}[\ell] = \delta[\ell]$$

Since the output-input cross-correlation can be written as:

$$r_{yx}[\ell] = h[\ell] * r_{xx}[\ell]$$

then, with $r_{xx}[\ell] = \delta[\ell]$, it follows:

$$r_{yx}[\ell] = h[\ell] * \delta[\ell] = h[\ell]$$

System identification

As the input or excitation process is WGN, then the output is WSS, and in many cases will be ergodic.

System identification

As the input or excitation process is WGN, then the output is WSS, and in many cases will be ergodic.

Hence, the cross-correlation (and therefore system impulse response) can be estimated from a single realisation using the *sample cross-correlation function*:

$$\hat{r}_{yx}[\ell] = \frac{1}{N} \sum_{n=0}^{N-1-|l|} y[n+|\ell|]\, x[n]\,, \quad |\ell| < N$$

$$\hat{r}'_{yx}[\ell] = \frac{1}{N-|l|} \sum_{n=0}^{N-1-|\ell|} y[n+|\ell|]\, x[n]\,, \quad |\ell| < N$$

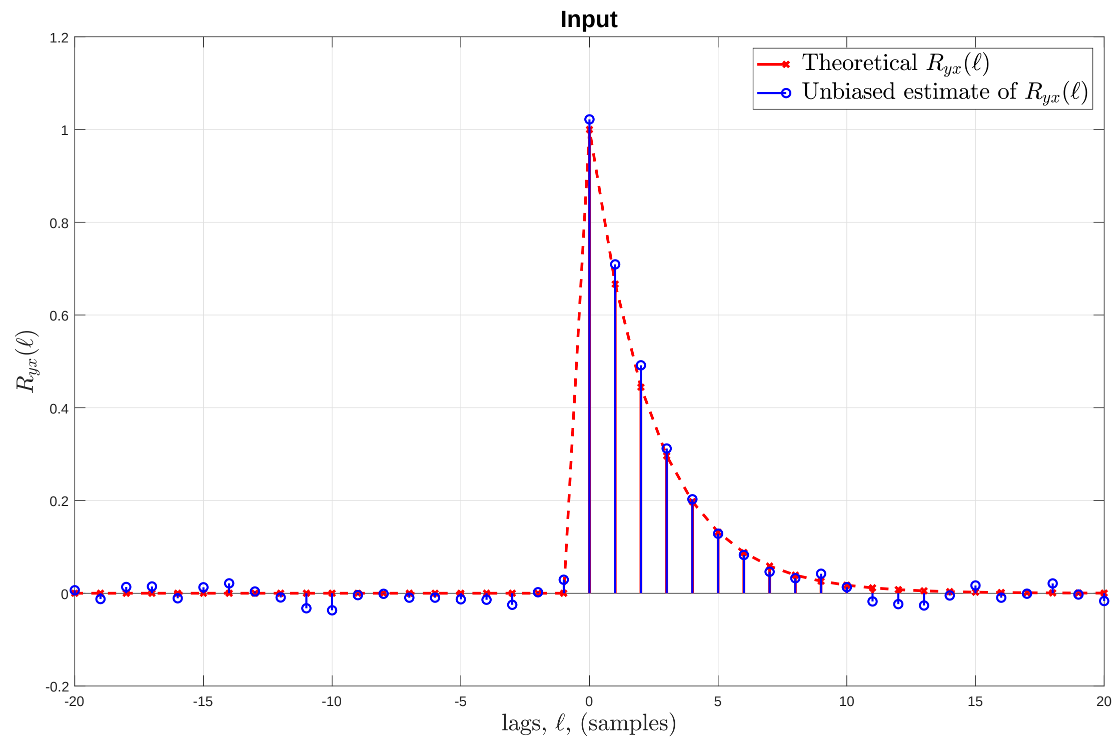
It is simple to generate an example in MATLAB.

System identification

Example (Low-pass filter). A system is described by $y[n] = \frac{2}{3} y[n-1] + x[n]$, although this is not known to the observer initially. By driving the system with WGN, calculate the impulse response of the system through numerical simulation.

System identification

Example (Low-pass filter). A system is described by $y[n] = \frac{2}{3} y[n-1] + x[n]$, although this is not known to the observer initially. By driving the system with WGN, calculate the impulse response of the system through numerical simulation.



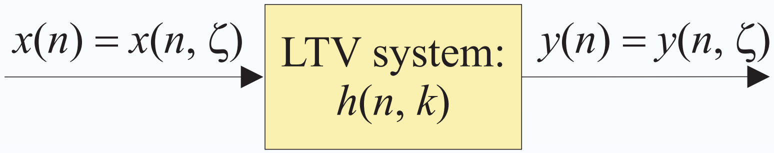
The theoretical impulse response $h[n] = \left(\frac{2}{3}\right)^n u[n]$ and the time-averaged estimate of the cross-correlation $\hat{R}_{yx}[\ell]$.

System identification

– End-of-Topic 58: **Application of Cross-Correlation to System Identification** –

Any Questions?

LTV Systems with Nonstationary Inputs



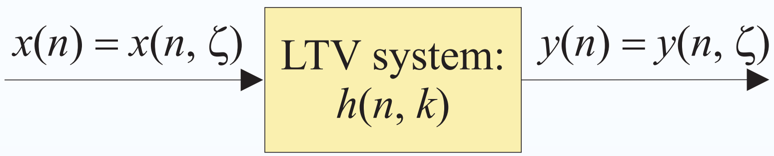
General LTV system with nonstationary input

The input and output are related by the generalised convolution:

$$y(n) = \sum_{k=-\infty}^{\infty} h(n, k)\, x(k)$$

where $h(n, k)$ is the response at time-index n to an impulse occurring at the system input at time-index k.

LTV Systems with Nonstationary Inputs



General LTV system with nonstationary input

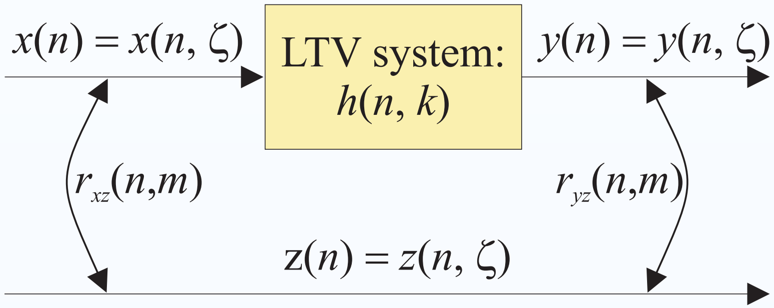
The input and output are related by the generalised convolution:

$$y(n) = \sum_{k=-\infty}^{\infty} h(n,k)\, x(k)$$

where $h(n,k)$ is the response at time-index n to an impulse occurring at the system input at time-index k.

- The mean, autocorrelation and autocovariance sequences of the output, $y(n)$, as well as the cross-correlation and cross-covariance functions between the input and the output, can be calculated in a similar way as for LTI systems with stationary inputs.

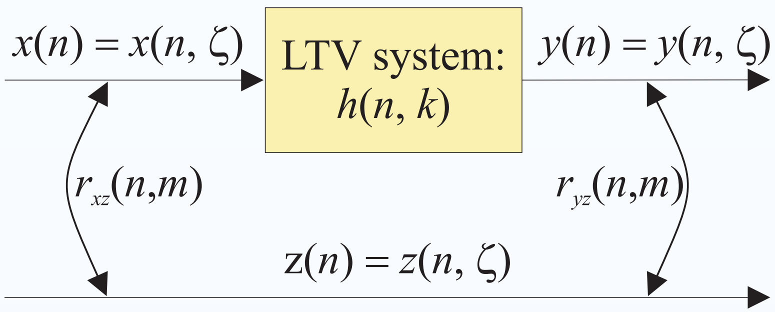
Linear Transformations on Cross-correlation



Cross-correlation with respect to a third random process.

- A random process $x[n]$ is transformed by a linear time-varying (LTV) system to produce another signal $y[n]$.
- The process $x[n]$ is related to a third process $z[n]$, and $r_{xz}[n_1, n_2]$ is known. It is desirable to find $r_{yz}[n_1, n_2]$.

Linear Transformations on Cross-correlation

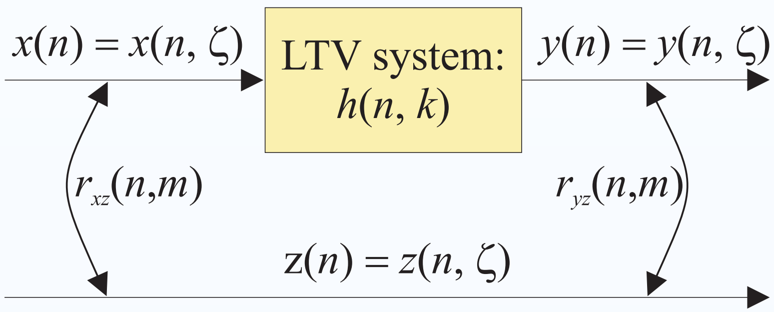


Cross-correlation with respect to a third random process.

- A random process $x[n]$ is transformed by a LTV system to produce another signal $y[n]$.
- The process $x[n]$ is related to a third process $z[n]$, and $r_{xz}[n_1, n_2]$ is known. It is desirable to find $r_{yz}[n_1, n_2]$.
- The response of the LTV system to $x[n]$ is:

$$y[n] = \sum_{k \in \mathbb{Z}} h[n, k] \; x[k]$$

Linear Transformations on Cross-correlation



Cross-correlation with respect to a third random process.

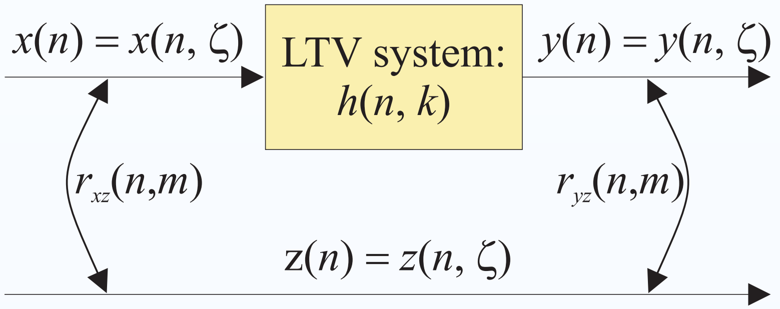
- A random process $x[n]$ is transformed by a LTV system to produce another signal $y[n]$.
- The response of the LTV system to $x[n]$ is:

$$y[n] = \sum_{k \in \mathbb{Z}} h[n, k] \; x[k]$$

Hence, multiplying both sides by $z^*[m]$ and taking expectations:

$$r_{yz}[n, m] = \sum_{k \in \mathbb{Z}} h[n, k] \; r_{xz}[k, m] = h[n, k] * r_{xz}[k, m]$$

Linear Transformations on Cross-correlation



- The response of the LTV system to $x[n]$ is:

$$y[n] = \sum_{k \in \mathbb{Z}} h[n, k] \; x[k]$$

Hence, multiplying both sides by $z^*[m]$ and taking expectations:

$$r_{yz}[n, m] = \sum_{k \in \mathbb{Z}} h[n, k] \; r_{xz}[k, m] = h[n, k] * r_{xz}[k, m]$$

If the system is LTI, then this simplifies to:

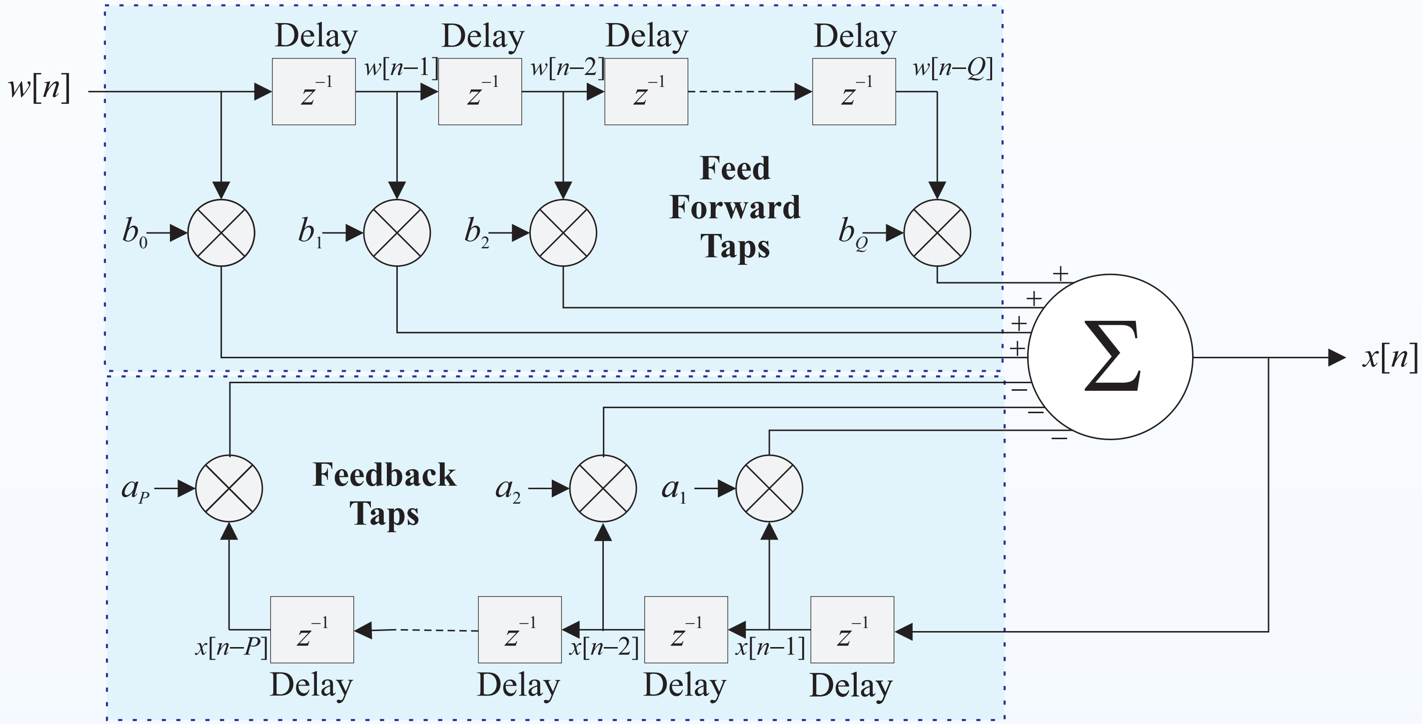
$$r_{yz}[\ell] = \sum_{k \in \mathbb{Z}} h[k] \; r_{xz}[\ell - k] = h[\ell] * r_{xz}[\ell]$$

Linear Transformations on Cross-correlation

– End-of-Topic 59: **Analysis of LTV systems and other special cases** –

Any Questions?

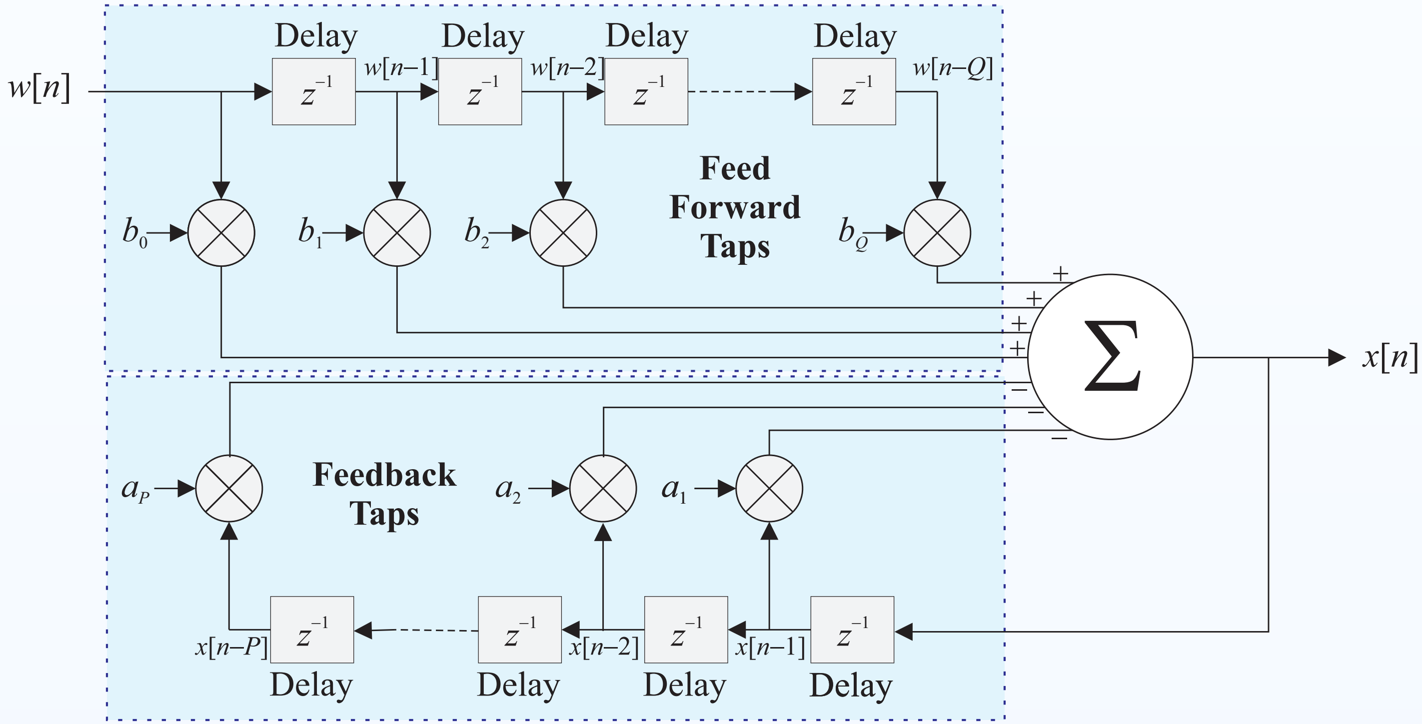
Analysis with Difference Equations



Difference-equation description of a LTI system.

- A mathematically elegant analysis of stochastic systems comes when a LTI system can be represented by difference equations.

Analysis with Difference Equations



Difference-equation description of a LTI system.

- A mathematically elegant analysis of stochastic systems comes when a LTI system can be represented by difference equations.
- The difference equation offers an alternative representation of the results that can sometimes be quite useful and important.

Analysis with Difference Equations

- A mathematically elegant analysis of stochastic systems comes when a LTI system can be represented by difference equations.
- The difference equation offers an alternative representation of the results that can sometimes be quite useful and important.
  - It is possible to use a combination of methods, such as taking the transfer function of a difference to find the impulse response, and then use convolution.
  - The purpose of the difference equation approach is to do the calculations in a single approach.

Analysis with Difference Equations

- A mathematically elegant analysis of stochastic systems comes when a LTI system can be represented by difference equations.
- The difference equation offers an alternative representation of the results that can sometimes be quite useful and important.
- Consider a LTI system that can be represented by:

$$y[n] = -\sum_{p=1}^{P} a_p\, y[n-p] + \sum_{q=0}^{Q} b_q\, x[n-q]$$

Analysis with Difference Equations

- A mathematically elegant analysis of stochastic systems comes when a LTI system can be represented by difference equations.
- The difference equation offers an alternative representation of the results that can sometimes be quite useful and important.
- Consider a LTI system that can be represented by:

$$y[n] = -\sum_{p=1}^{P} a_p \, y[n-p] + \sum_{q=0}^{Q} b_q \, x[n-q]$$

- Assuming that both $x[n]$ and $y[n]$ are stationary processes, then taking expectations of both sides gives:

$$\mu_y = \frac{\sum_{q=0}^{Q} b_q}{1 + \sum_{p=1}^{P} a_p} \mu_x$$

Analysis with Difference Equations

Assuming stationarity, then multiplying the system equation throughout by $y^*[n-\ell]$ and taking expectations gives:

$$\sum_{p=0}^{P} a_p \, r_{yy}[\ell - p] = \sum_{q=0}^{Q} b_q \, r_{xy}[\ell - q]$$

where $a_0 \triangleq 1$.

Analysis with Difference Equations

Assuming stationarity, then multiplying the system equation throughout by $y^*[n-\ell]$ and taking expectations gives:

$$\sum_{p=0}^{P} a_p \, r_{yy}[\ell - p] = \sum_{q=0}^{Q} b_q \, r_{xy}[\ell - q]$$

Similarly, instead multiply though by $x^*[n-\ell]$ to give:

$$\sum_{p=0}^{P} a_p \, r_{yx}[\ell - p] = \sum_{q=0}^{Q} b_q \, r_{xx}[\ell - q]$$

These equations may be used to solve for $r_{yy}[\ell]$ and $r_{xy}[\ell]$.

Analysis with Difference Equations

Assuming stationarity, then multiplying the system equation throughout by $y^*[n-\ell]$ and taking expectations gives:

$$\sum_{p=0}^{P} a_p\, r_{yy}[\ell-p] = \sum_{q=0}^{Q} b_q\, r_{xy}[\ell-q]$$

Similarly, instead multiply though by $x^*[n-\ell]$ to give:

$$\sum_{p=0}^{P} a_p\, r_{yx}[\ell-p] = \sum_{q=0}^{Q} b_q\, r_{xx}[\ell-q]$$

These equations may be used to solve for $r_{yy}[\ell]$ and $r_{xy}[\ell]$.

- Note the statistics auto- and cross-correlation statistics satisfy the original difference equations.

Analysis with Difference Equations

Assuming stationarity, then multiplying the system equation throughout by $y^*[n-\ell]$ and taking expectations gives:

$$\sum_{p=0}^{P} a_p\, r_{yy}[\ell - p] = \sum_{q=0}^{Q} b_q\, r_{xy}[\ell - q]$$

Similarly, instead multiply though by $x^*[n-\ell]$ to give:

$$\sum_{p=0}^{P} a_p\, r_{yx}[\ell - p] = \sum_{q=0}^{Q} b_q\, r_{xx}[\ell - q]$$

These equations may be used to solve for $r_{yy}[\ell]$ and $r_{xy}[\ell]$.

- Note the statistics auto- and cross-correlation statistics satisfy the original difference equations.
- Similar expressions can be obtained for the covariance sequences.

Analysis with Difference Equations

Example ([Manolakis:2000, Example 3.6.2, Page 141]). Let $x[n]$ be generated by the first order difference equation given by:

$$x[n] = \alpha\, x[n-1] + w[n]\,, \quad |\alpha| \le 1, n \in \mathbb{Z} \qquad \bowtie$$

where $w[n] \sim \mathcal{N}\left(\mu_w,\, \sigma_w^2\right)$ is an i. i. d. WGN process.

- Demonstrate that $x[n]$ is stationary, and calculate μ_x.
- Determine the autocovariance and autocorrelation sequences, $\gamma_{xx}[\ell]$ and $r_{xx}[\ell]$.

Analysis with Difference Equations

Example ([Manolakis:2000, Example 3.6.2, Page 141]). Let $x[n]$ be generated by the first order difference equation given by:

$$x[n] = \alpha\, x[n-1] + w[n]\,, \quad |\alpha| \leq 1, n \in \mathbb{Z}$$

where $w[n] \sim \mathcal{N}\left(\mu_w,\, \sigma_w^2\right)$ is an i. i. d. WGN process.

- Demonstrate that $x[n]$ is stationary, and calculate μ_x.
- Determine the autocovariance and autocorrelation sequences, $\gamma_{xx}[\ell]$ and $r_{xx}[\ell]$.

SOLUTION. The output of a LTI system with a stationary input is always stationary. □

Analysis with Difference Equations

Example ([Manolakis:2000, Example 3.6.2, Page 141]). Let $x[n]$ be generated by the first order difference equation given by:

$$x[n] = \alpha\, x[n-1] + w[n]\,, \quad |\alpha| \leq 1, n \in \mathbb{Z}$$

where $w[n] \sim \mathcal{N}\left(\mu_w, \sigma_w^2\right)$ is an i. i. d. WGN process.

- Demonstrate that $x[n]$ is stationary, and calculate μ_x.
- Determine the autocovariance and autocorrelation sequences, $\gamma_{xx}[\ell]$ and $r_{xx}[\ell]$.

SOLUTION. The output of a LTI system with a stationary input is always stationary.

- It follows directly from the results above that:

$$\mu_x = \frac{\mu_w}{1-\alpha}$$

□

Analysis with Difference Equations

Example ([Manolakis:2000, Example 3.6.2, Page 141]). Let $x[n]$ be generated by the first order difference equation given by:

$$x[n] = \alpha\, x[n-1] + w[n]\,, \quad |\alpha| \leq 1, n \in \mathbb{Z}$$

- Determine $\gamma_{xx}[\ell]$ and $r_{xx}[\ell]$.

SOLUTION. - Using the results for the input-output covariance:

$$\gamma_{xx}[\ell] - \alpha\, \gamma_{xx}[\ell-1] = \gamma_{wx}[\ell]$$
$$\gamma_{xw}[\ell] - \alpha\, \gamma_{xw}[\ell-1] = \gamma_{ww}[\ell]$$

□

Analysis with Difference Equations

Example ([Manolakis:2000, Example 3.6.2, Page 141]). Let $x[n]$ be generated by the first order difference equation given by:

$$x[n] = \alpha\, x[n-1] + w[n]\,, \quad |\alpha| \leq 1, n \in \mathbb{Z}$$

- Determine $\gamma_{xx}[\ell]$ and $r_{xx}[\ell]$.

SOLUTION. - Using the results for the input-output covariance:

$$\gamma_{xx}[\ell] - \alpha\, \gamma_{xx}[\ell - 1] = \gamma_{wx}[\ell]$$
$$\gamma_{xw}[\ell] - \alpha\, \gamma_{xw}[\ell - 1] = \gamma_{ww}[\ell]$$

- $x[n]$ cannot depend on future $w[n] \Rightarrow \gamma_{xw}[\ell] = 0,\ \ell < 0.$ □

Analysis with Difference Equations

Example ([Manolakis:2000, Example 3.6.2, Page 141]). Let $x[n]$ be generated by the first order difference equation given by:

$$x[n] = \alpha\, x[n-1] + w[n]\,, \quad |\alpha| \leq 1, n \in \mathbb{Z}$$

- Determine $\gamma_{xx}[\ell]$ and $r_{xx}[\ell]$.

SOLUTION. Using the results for the input-output covariance:

$$\gamma_{xx}[\ell] - \alpha\, \gamma_{xx}[\ell - 1] = \gamma_{wx}[\ell]$$
$$\gamma_{xw}[\ell] - \alpha\, \gamma_{xw}[\ell - 1] = \gamma_{ww}[\ell]$$

- $x[n]$ cannot depend on future $w[n] \Rightarrow \gamma_{xw}[\ell] = 0,\ \ell < 0$.
  - This is shown by evaluating $r_{xw}[\ell] = \mathbb{E}\left[x[n]\ w^*[n-\ell]\right]$, and noting that $x[n]$ and $w[n]$ are independent. □

Analysis with Difference Equations

Example ([Manolakis:2000, Example 3.6.2, Page 141]). Let $x[n]$ be generated by the first order difference equation given by:

$$x[n] = \alpha\, x[n-1] + w[n]\,, \quad |\alpha| \leq 1, n \in \mathbb{Z}$$

- Determine $\gamma_{xx}[\ell]$ and $r_{xx}[\ell]$.

SOLUTION. Using the results for the input-output covariance:

$$\gamma_{xx}[\ell] - \alpha\, \gamma_{xx}[\ell-1] = \gamma_{wx}[\ell]$$
$$\gamma_{xw}[\ell] - \alpha\, \gamma_{xw}[\ell-1] = \gamma_{ww}[\ell]$$

- $x[n]$ cannot depend on future $w[n] \Rightarrow \gamma_{xw}[\ell] = 0,\ \ell < 0$.
  - If $\ell < 0$, then $w[n-\ell]$ is a sample with time-index greater than that of $x[n]$, or in otherwords a future value. □

Analysis with Difference Equations

Example ([Manolakis:2000, Example 3.6.2, Page 141]). Let $x[n]$ be generated by the first order difference equation given by:

$$x[n] = \alpha\, x[n-1] + w[n]\,, \quad |\alpha| \le 1, n \in \mathbb{Z}$$

- Determine $\gamma_{xx}[\ell]$ and $r_{xx}[\ell]$.

SOLUTION. Since $\gamma_{ww}[\ell] = \sigma_w^2\, \delta[\ell]$, the second of the difference equations above becomes:

$$\gamma_{xw}[\ell] = \begin{cases} \alpha\, \gamma_{xw}[\ell - 1] & \ell > 0 \\ \sigma_w^2 & \ell = 0 \\ 0 & \ell < 0 \end{cases}$$

□

Analysis with Difference Equations

Example ([Manolakis:2000, Example 3.6.2, Page 141]). Let $x[n]$ be generated by the first order difference equation given by:

$$x[n] = \alpha\, x[n-1] + w[n]\,, \quad |\alpha| \leq 1, n \in \mathbb{Z}$$

- Determine $\gamma_{xx}[\ell]$ and $r_{xx}[\ell]$.

SOLUTION. Since $\gamma_{ww}[\ell] = \sigma_w^2\, \delta[\ell]$, the second of the difference equations above becomes:

$$\gamma_{xw}[\ell] = \begin{cases} \alpha\, \gamma_{xw}[\ell - 1] & \ell > 0 \\ \sigma_w^2 & \ell = 0 \\ 0 & \ell < 0 \end{cases} \qquad \square$$

Solving for $\ell \geq 0$ gives by repeated substitution, $\gamma_{xw}[\ell] = \alpha^\ell\, \sigma_w^2$, and zero for $\ell < 0$.

Analysis with Difference Equations

Example ([Manolakis:2000, Example 3.6.2, Page 141]). Let $x[n]$ be generated by the first order difference equation given by:

$$x[n] = \alpha\, x[n-1] + w[n]\,, \quad |\alpha| \leq 1,\, n \in \mathbb{Z}$$

- Determine $\gamma_{xx}[\ell]$ and $r_{xx}[\ell]$.

SOLUTION. • Since $\gamma_{wx}[\ell] = \gamma^*_{xw}[-\ell]$, then the difference equation for the autocovariance function of $x[n]$ simplifies to:

$$\gamma_{xx}[\ell] - \alpha\,\gamma_{xx}[\ell-1] = \begin{cases} 0 & \ell > 0 \\ \alpha^{-\ell}\,\sigma_w^2 & \ell \leq 0 \end{cases}$$

□

Analysis with Difference Equations

Example ([Manolakis:2000, Example 3.6.2, Page 141]). Let $x[n]$ be generated by the first order difference equation given by:

$$x[n] = \alpha\, x[n-1] + w[n]\,, \quad |\alpha| \leq 1, n \in \mathbb{Z}$$

- Determine $\gamma_{xx}[\ell]$ and $r_{xx}[\ell]$.

SOLUTION.
- Since $\gamma_{wx}[\ell] = \gamma^*_{xw}[-\ell]$, then :

$$\gamma_{xx}[\ell] - \alpha\,\gamma_{xx}[\ell-1] = \begin{cases} 0 & \ell > 0 \\ \alpha^{-\ell}\,\sigma_w^2 & \ell \leq 0 \end{cases}$$

- Note the solution for $\ell > 0$ is the solution of the homogeneous equation. □

Analysis with Difference Equations

Example ([Manolakis:2000, Example 3.6.2, Page 141]). Let $x[n]$ be generated by the first order difference equation given by:

$$x[n] = \alpha\, x[n-1] + w[n]\,, \quad |\alpha| \leq 1, n \in \mathbb{Z}$$

- Determine $\gamma_{xx}[\ell]$ and $r_{xx}[\ell]$.

SOLUTION.
- Since $\gamma_{wx}[\ell] = \gamma^*_{xw}[-\ell]$, then :

$$\gamma_{xx}[\ell] - \alpha\,\gamma_{xx}[\ell-1] = \begin{cases} 0 & \ell > 0 \\ \alpha^{-\ell}\,\sigma_w^2 & \ell \leq 0 \end{cases}$$

- Hence, since $\gamma_{xx}[\ell] = \gamma_{xx}[-\ell]$ for a real process, then this equation is solved by assuming the solution:

$$\gamma_{xx}[\ell] = a\,\alpha^{|\ell|} + b$$

Analysis with Difference Equations

Example ([Manolakis:2000, Example 3.6.2, Page 141]). Let $x[n]$ be generated by the first order difference equation given by:

$$x[n] = \alpha\, x[n-1] + w[n]\,, \quad |\alpha| \leq 1, n \in \mathbb{Z}$$

- Determine $\gamma_{xx}[\ell]$ and $r_{xx}[\ell]$.

SOLUTION. Assuming the solution:

$$\gamma_{xx}[\ell] = a\, \alpha^{|\ell|} + b$$

- a and b can be found by substituting the proposed solution for $\ell \leq 0$ into the difference equation:

$$a\, \alpha^{-\ell} + b - \alpha \left(a\, \alpha^{-(\ell-1)} + b \right) = \alpha^{-\ell}\, \sigma_w^2$$

$$\alpha^{-\ell} \left(1 - \alpha^2\right) a + (1-\alpha)\, b = \alpha^{-\ell}\, \sigma_w^2$$

Analysis with Difference Equations

Example ([Manolakis:2000, Example 3.6.2, Page 141]). Let $x[n]$ be generated by the first order difference equation given by:

$$x[n] = \alpha\, x[n-1] + w[n]\,, \quad |\alpha| \leq 1, n \in \mathbb{Z}$$

- Determine $\gamma_{xx}[\ell]$ and $r_{xx}[\ell]$.

SOLUTION. Assuming the solution:

$$\gamma_{xx}[\ell] = a\,\alpha^{|\ell|} + b$$

- from which it directly follows that $b = 0$ and $a = \sigma_x^2 = \frac{\sigma_w^2}{1-\alpha^2}$, corresponding to the case when $\ell = 0$. □

Analysis with Difference Equations

Example ([Manolakis:2000, Example 3.6.2, Page 141]). Let $x[n]$ be generated by the first order difference equation given by:

$$x[n] = \alpha\, x[n-1] + w[n]\,, \quad |\alpha| \le 1,\, n \in \mathbb{Z}$$

- Determine $\gamma_{xx}[\ell]$ and $r_{xx}[\ell]$.

SOLUTION. • Hence, in conclusion

$$\gamma_{xx}[\ell] = \frac{\sigma_w^2}{1-\alpha^2}\, \alpha^{|\ell|}$$

□

Analysis with Difference Equations

Example ([Manolakis:2000, Example 3.6.2, Page 141]). Let $x[n]$ be generated by the first order difference equation given by:

$$x[n] = \alpha\, x[n-1] + w[n]\,, \quad |\alpha| \le 1, n \in \mathbb{Z}$$

- Determine $\gamma_{xx}[\ell]$ and $r_{xx}[\ell]$.

SOLUTION. • Hence, in conclusion

$$\gamma_{xx}[\ell] = \frac{\sigma_w^2}{1-\alpha^2}\, \alpha^{|\ell|}$$

Using the relationship that $r_{xx}[\ell] = \gamma_{xx}[\ell] + \mu_x^2$, it follows that the output auto-correlation is given by:

$$r_{xx}[\ell] = \frac{\sigma_w^2}{1-\alpha^2}\, \alpha^{|\ell|} + \frac{\mu_w^2}{(1-\alpha)^2} \qquad \square$$

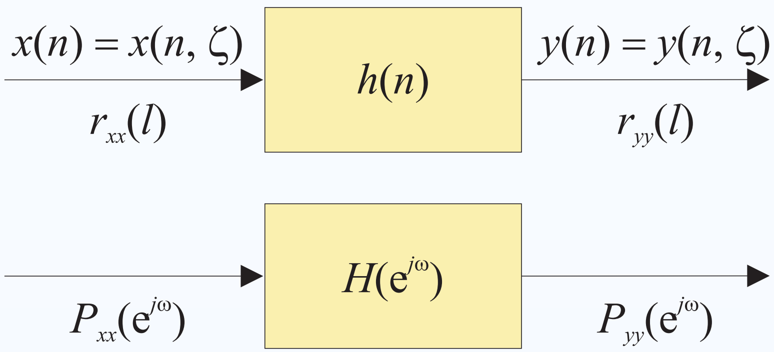
Analysis with Difference Equations

– End-of-Topic 60: **Analysis of input-output statistics using difference equation approach** –

Any Questions?

Frequency-Domain Analysis of LTI systems

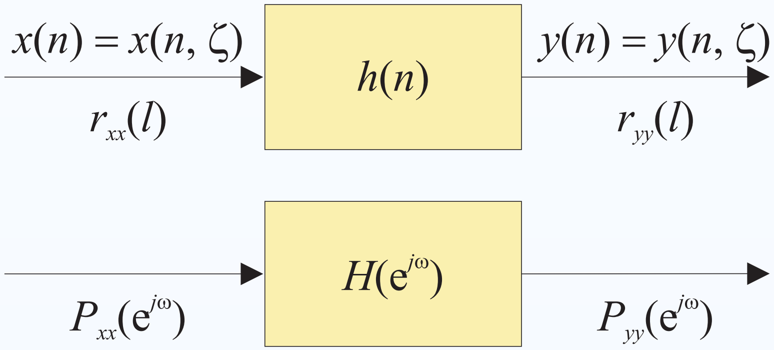
Now consider how a LTI transformation affects the power spectra and complex spectra of a stationary random process.



LTI system with WSS input.

Frequency-Domain Analysis of LTI systems

Now consider how a LTI transformation affects the power spectra and complex spectra of a stationary random process.

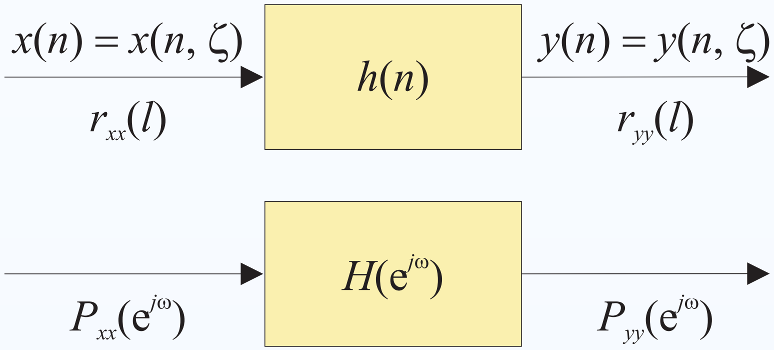


LTI system with WSS input.

Taking the DTFT of the time-domain relationships for the input-output statistics in terms of the system impulse response leads to the following spectral densities:

Frequency-Domain Analysis of LTI systems

Now consider how a LTI transformation affects the power spectra and complex spectra of a stationary random process.

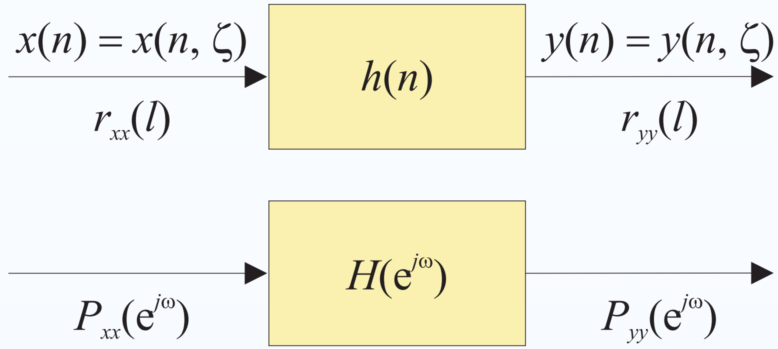


LTI system with WSS input.

Taking the DTFT of the time-domain relationships for the input-output statistics in terms of the system impulse response leads to the following spectral densities:

$$
\begin{aligned}
r_{xy}[\ell] &= h^*[-\ell] * r_{xx}[\ell] &\Rightarrow\quad P_{xy}(e^{j\omega}) &= H^*(e^{j\omega})\, P_{xx}(e^{j\omega}) \\
r_{yx}[\ell] &= h[\ell] * r_{xx}[\ell] &\Rightarrow\quad P_{yx}(e^{j\omega}) &= H(e^{j\omega})\, P_{xx}(e^{j\omega}) \\
r_{yy}[\ell] &= h^*[-\ell] * h[\ell] * r_{xx}[\ell] &\Rightarrow\quad P_{yy}(e^{j\omega}) &= |H(e^{j\omega})|^2\, P_{xx}(e^{j\omega})
\end{aligned}
$$

Frequency-Domain Analysis of LTI systems



Taking the DTFT of the time-domain relationships :

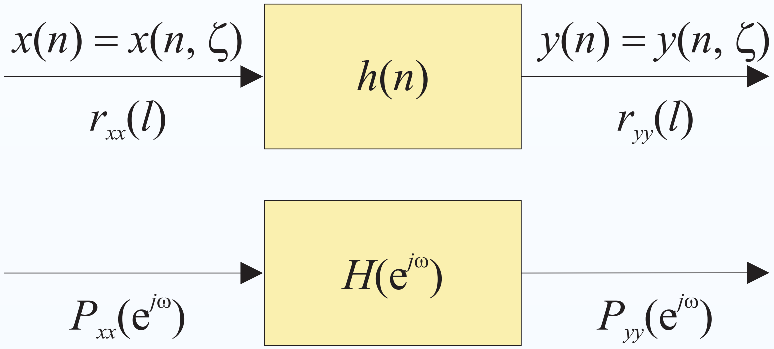
$$r_{xy}[\ell] = h^*[-\ell] * r_{xx}[\ell] \quad \Rightarrow \quad P_{xy}(e^{j\omega}) = H^*(e^{j\omega})\, P_{xx}(e^{j\omega})$$

$$r_{yx}[\ell] = h[\ell] * r_{xx}[\ell] \quad \Rightarrow \quad P_{yx}(e^{j\omega}) = H(e^{j\omega})\, P_{xx}(e^{j\omega})$$

$$r_{yy}[\ell] = h^*[-\ell] * h[\ell] * r_{xx}[\ell] \quad \Rightarrow \quad P_{yy}(e^{j\omega}) = |H(e^{j\omega})|^2\, P_{xx}(e^{j\omega})$$

- If the input and output autocorrelations or autospectral densities are known, the magnitude response of a system $|H(e^{j\omega})|$ can be determined, but not the phase response.

Frequency-Domain Analysis of LTI systems



Taking the DTFT of the time-domain relationships :

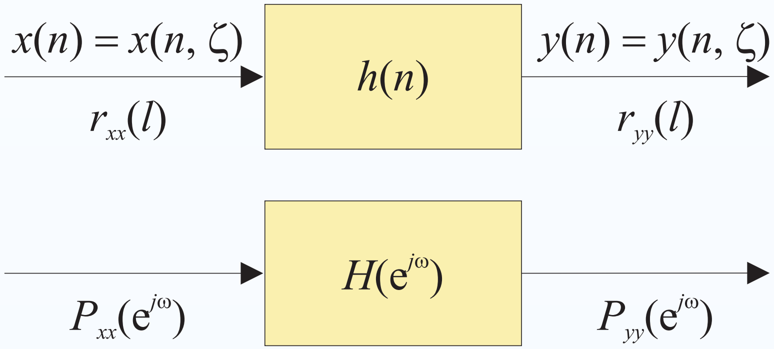
$$r_{xy}[\ell] = h^*[-\ell] * r_{xx}[\ell] \quad \Rightarrow \quad P_{xy}(e^{j\omega}) = H^*(e^{j\omega})\, P_{xx}(e^{j\omega})$$

$$r_{yx}[\ell] = h[\ell] * r_{xx}[\ell] \quad \Rightarrow \quad P_{yx}(e^{j\omega}) = H(e^{j\omega})\, P_{xx}(e^{j\omega})$$

$$r_{yy}[\ell] = h^*[-\ell] * h[\ell] * r_{xx}[\ell] \quad \Rightarrow \quad P_{yy}(e^{j\omega}) = |H(e^{j\omega})|^2\, P_{xx}(e^{j\omega})$$

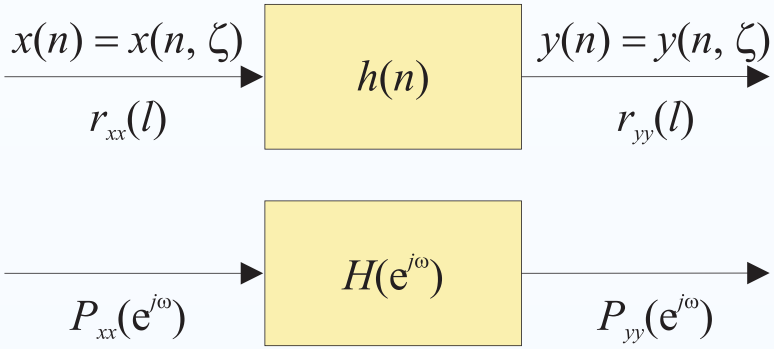
- If the input and output autocorrelations or autospectral densities are known, the magnitude response of a system $|H(e^{j\omega})|$ can be determined, but not the phase response.
- Only cross-spectral information can help determine phase.

Frequency-Domain Analysis of LTI systems



- A set of similar relations can be derived for the complex spectral density function.

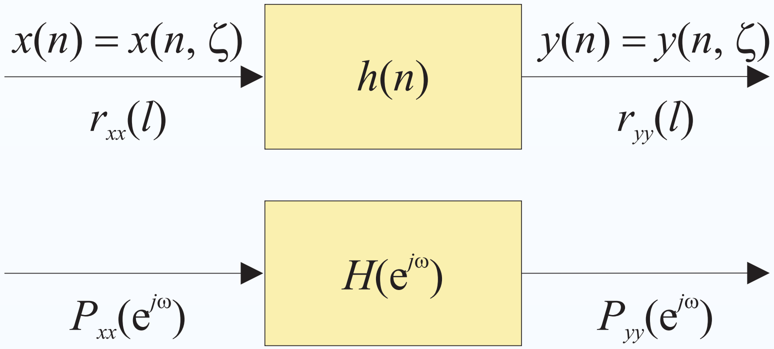
Frequency-Domain Analysis of LTI systems



- Specifically, if: $h[\ell] \overset{z}{\rightleftharpoons} H(z)$, then:

$$h^*[-\ell] \overset{z}{\rightleftharpoons} H^*\left({}^1/_{z^*}\right)$$

Frequency-Domain Analysis of LTI systems



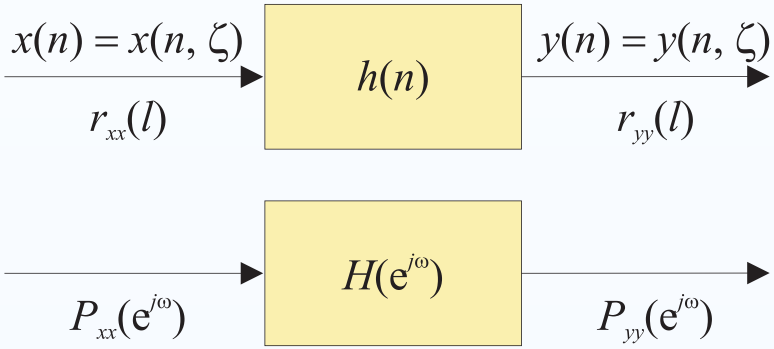
- Specifically, if: $h[\ell] \overset{z}{\rightleftharpoons} H(z)$, then:

$$h^*[-\ell] \overset{z}{\rightleftharpoons} H^*\left({}^1/_{z^*}\right)$$

Therefore, the input output relationships:

$$\begin{aligned} r_{xy}[\ell] &= h^*[-\ell] * r_{xx}[\ell] \\ r_{yx}[\ell] &= h[\ell] * r_{xx}[\ell] \\ r_{yy}[\ell] &= h[\ell] * r_{xy}[\ell] \\ &= h[\ell] * h^*[-\ell] * r_{xx}[\ell] \end{aligned}$$

Frequency-Domain Analysis of LTI systems



- Specifically, if: $h[\ell] \overset{z}{\rightleftharpoons} H(z)$, then:

$$h^*[-\ell] \overset{z}{\rightleftharpoons} H^*\left({}^1/_{z^*}\right)$$

Transform to the spectral relationships:

$$P_{xy}(z) = H^*\left({}^1/_{z^*}\right) P_{xx}(z)$$

$$P_{yx}(z) = H(z)\, P_{xx}(z)$$

$$P_{yy}(z) = H(z)\, P_{xy}(z)$$

$$P_{yy}(z) = H(z)\, H^*\left({}^1/_{z^*}\right) P_{xx}(z)$$

Frequency-Domain Analysis of LTI systems

$$P_{xy}(z) = H^*\left({}^1/_{z^*}\right)\, P_{xx}(z)$$
$$P_{yx}(z) = H(z)\, P_{xx}(z)$$
$$P_{yy}(z) = H(z)\, P_{xy}(z)$$
$$P_{yy}(z) = H(z)\, H^*\left({}^1/_{z^*}\right)\, P_{xx}(z)$$

- Note that $P_{yy}(z)$ satisfies the required property for a complex spectral density function, namely that $P_{yy}(z) = P_{yy}^*\left({}^1/_{z^*}\right)$.

Frequency-Domain Analysis of LTI systems

$$P_{xy}(z) = H^*\left(1/z^*\right) P_{xx}(z)$$
$$P_{yx}(z) = H(z)\, P_{xx}(z)$$
$$P_{yy}(z) = H(z)\, P_{xy}(z)$$
$$P_{yy}(z) = H(z)\, H^*\left(1/z^*\right) P_{xx}(z)$$

- Note that $P_{yy}(z)$ satisfies the required property for a complex spectral density function, namely that $P_{yy}(z) = P^*_{yy}\left(1/z^*\right)$.
- Also, note the following result for real filters that make the above equations simplify accordingly.

Frequency-Domain Analysis of LTI systems

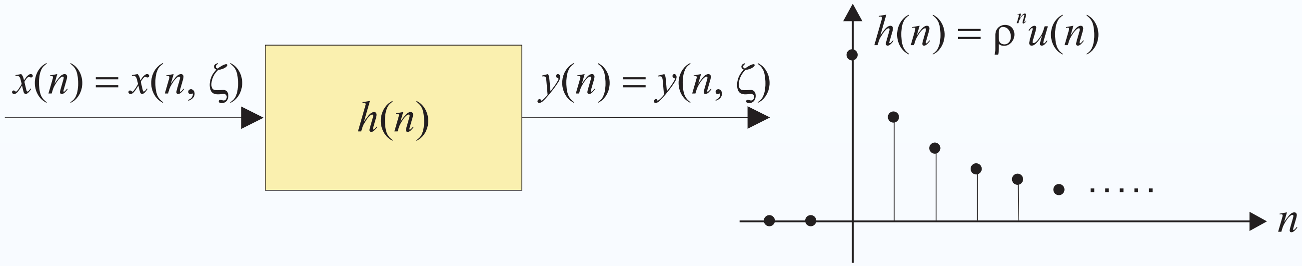
$$
\begin{aligned}
P_{xy}(z) &= H^*\left(1/z^*\right)\, P_{xx}(z) \\
P_{yx}(z) &= H(z)\, P_{xx}(z) \\
P_{yy}(z) &= H(z)\, P_{xy}(z) \\
P_{yy}(z) &= H(z)\, H^*\left(1/z^*\right)\, P_{xx}(z)
\end{aligned}
$$

- Note that $P_{yy}(z)$ satisfies the required property for a complex spectral density function, namely that $P_{yy}(z) = P_{yy}^*\left(1/z^*\right)$.
- Also, note the following result for real filters that make the above equations simplify accordingly.

Theorem (Transfer function for a real filter). For a real filter:

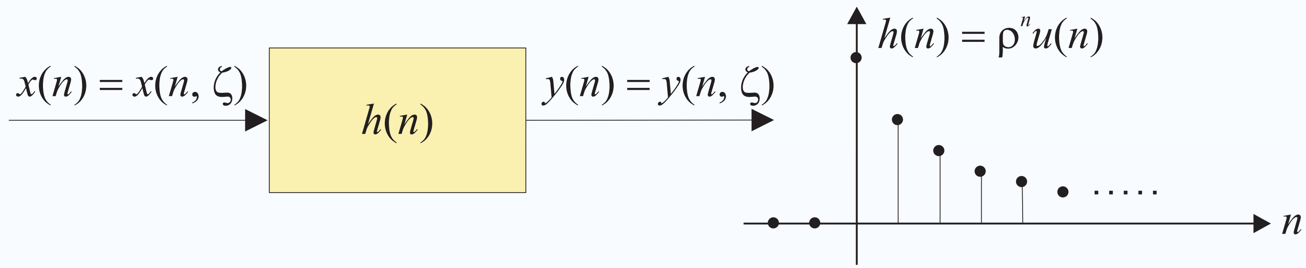
$$
h[-\ell] \overset{z}{\rightleftharpoons} H^*\left(\frac{1}{z^*}\right) = H(z^{-1}) \qquad \diamond
$$

Frequency-Domain Analysis of LTI systems



Example (Simple Example:). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2\,\delta[\ell]$. Calculate the PSD, CPSD and the complex spectral densities.

Frequency-Domain Analysis of LTI systems



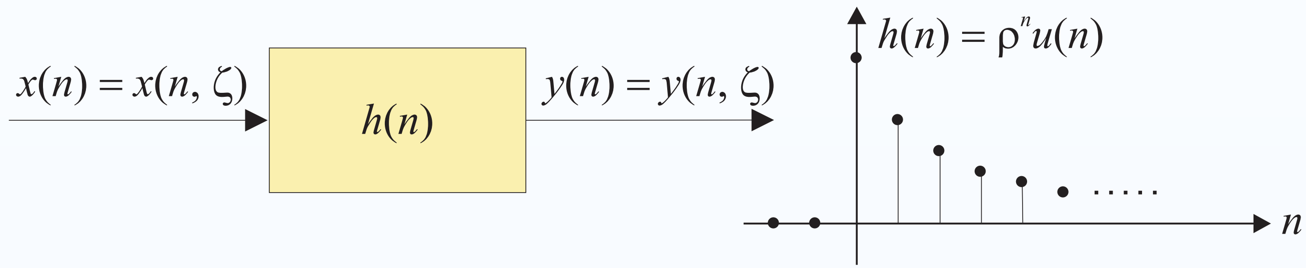
Example (Simple Example:). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2\,\delta[\ell]$. Calculate the PSD, CPSD and the complex spectral densities.

SOLUTION. The impulse response $h[n] = \rho^n u[n]$ has system transfer function:

$$H(z) = \frac{1}{1 - \rho\, z^{-1}}$$

□

Frequency-Domain Analysis of LTI systems



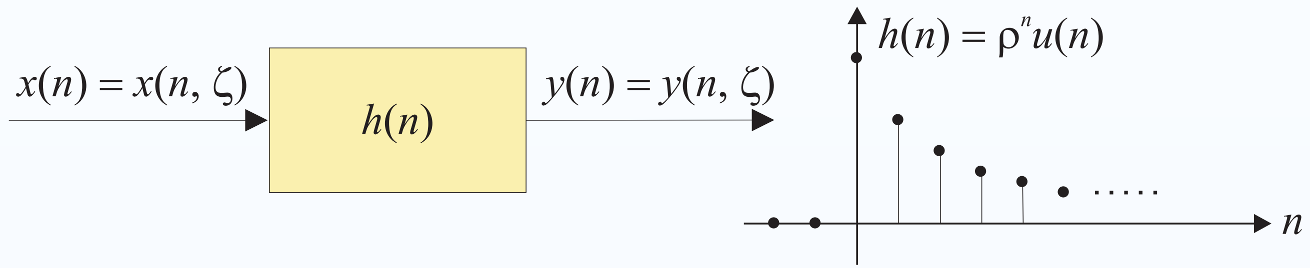
Example (Simple Example:). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2\, \delta[\ell]$. Calculate the PSD, CPSD and the complex spectral densities.

SOLUTION. Since $\gamma_{xx}[\ell] = \sigma_x^2 \delta[\ell]$, then:

$$r_{xx}[\ell] = \gamma_{xx}[\ell] + \mu_x^2 = \sigma_x^2 \delta[\ell] + |\mu_x|^2$$

□

Frequency-Domain Analysis of LTI systems



Example (Simple Example:). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2\,\delta[\ell]$. Calculate the PSD, CPSD and the complex spectral densities.

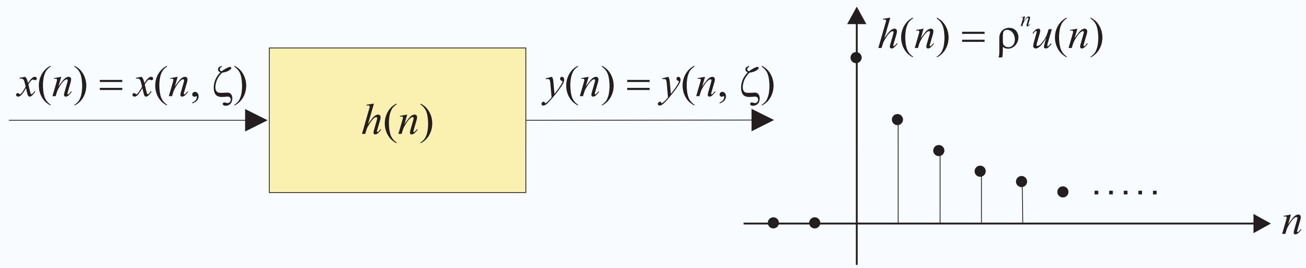
SOLUTION. Since $\gamma_{xx}[\ell] = \sigma_x^2\delta[\ell]$, then:

$$r_{xx}[\ell] = \gamma_{xx}[\ell] + \mu_x^2 = \sigma_x^2\delta[\ell] + |\mu_x|^2$$

Taking z-transforms gives:

$$\begin{aligned} P_{xx}(z) &= \sigma_x^2 + 2\pi|\mu_x|^2\delta(z - e^{j0}) \\ &= \sigma_x^2 + 2\pi|\mu_x|^2\delta(z - 1) \end{aligned}$$

Frequency-Domain Analysis of LTI systems



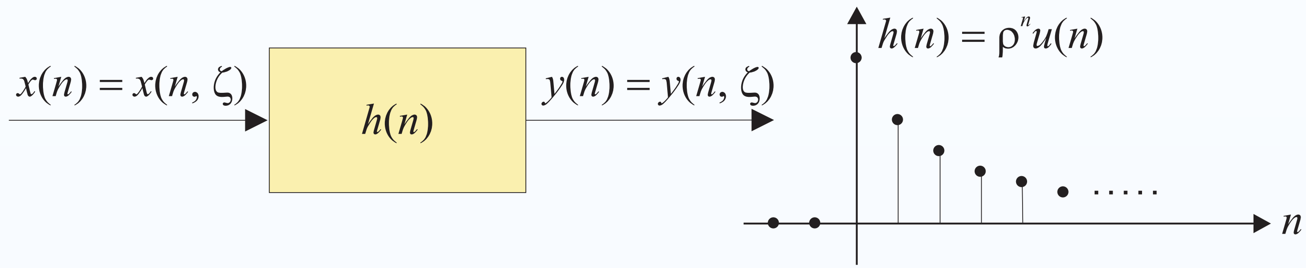
Example (Simple Example:). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2\, \delta[\ell]$. Calculate the PSD, CPSD and the complex spectral densities.

SOLUTION. Hence, the complex cross-spectral density is:

$$P_{xy}(z) = H^*\left({}^1\!/_{z^*}\right)\, P_{xx}(z)$$

□

Frequency-Domain Analysis of LTI systems



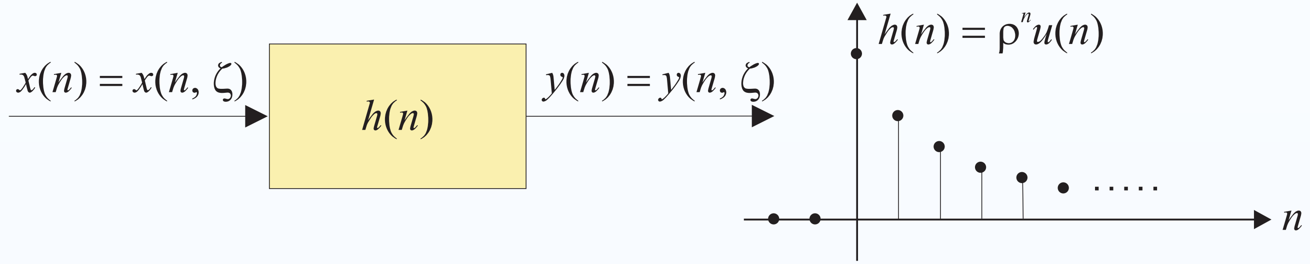
Example (Simple Example:). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2\, \delta[\ell]$. Calculate the PSD, CPSD and the complex spectral densities.

SOLUTION. Hence, the complex cross-spectral density is:

$$P_{xy}(z) = H^*\left({}^{1}\!/_{z^*}\right)\, P_{xx}(z)$$

$$= \left(\frac{1}{1 - \rho\left[\frac{1}{z^*}\right]^{-1}}\right)^* \left[\sigma_x^2 + 2\pi |\mu_x|^2 \delta(z-1)\right]$$

Frequency-Domain Analysis of LTI systems



Example (Simple Example:). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2\,\delta[\ell]$. Calculate the PSD, CPSD and the complex spectral densities.

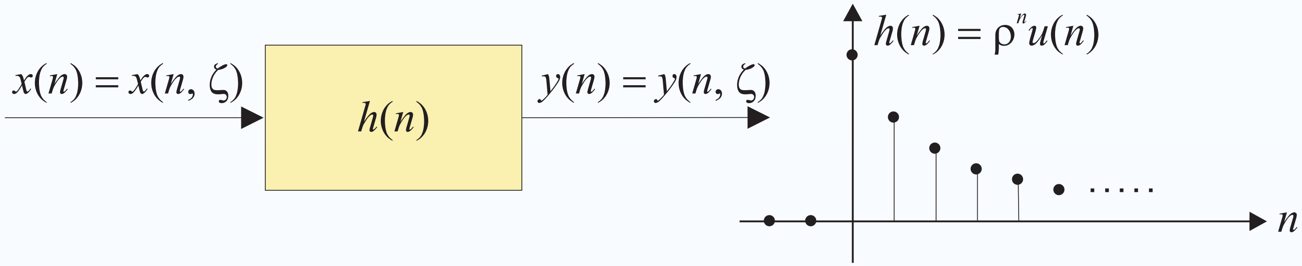
SOLUTION. Hence, the complex cross-spectral density is:

$$P_{xy}(z) = H^*\left({}^1/_{z^*}\right)\, P_{xx}(z)$$

$$= \left(\frac{1}{1-\rho\left[\frac{1}{z^*}\right]^{-1}}\right)^* \left[\sigma_x^2 + 2\pi|\mu_x|^2\delta(z-1)\right]$$

$$= \frac{\sigma_x^2}{1-\rho^* z} + \frac{2\pi|\mu_x|^2}{1-\rho^* z}\delta(z-1)$$

Frequency-Domain Analysis of LTI systems



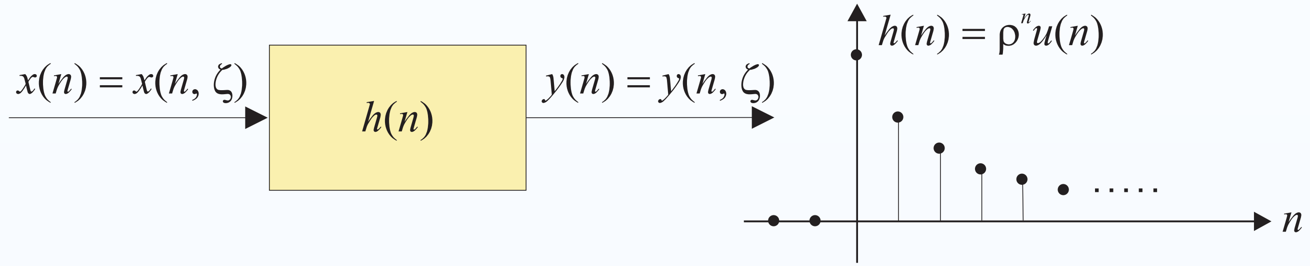
Example (Simple Example:). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2\,\delta[\ell]$. Calculate the PSD, CPSD and the complex spectral densities.

SOLUTION. Moreover, the complex spectral density is given by:

$$P_{yy}(z) = H(z)\;P_{xy}(z)$$

□

Frequency-Domain Analysis of LTI systems



Example (Simple Example:). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2\,\delta[\ell]$. Calculate the PSD, CPSD and the complex spectral densities.

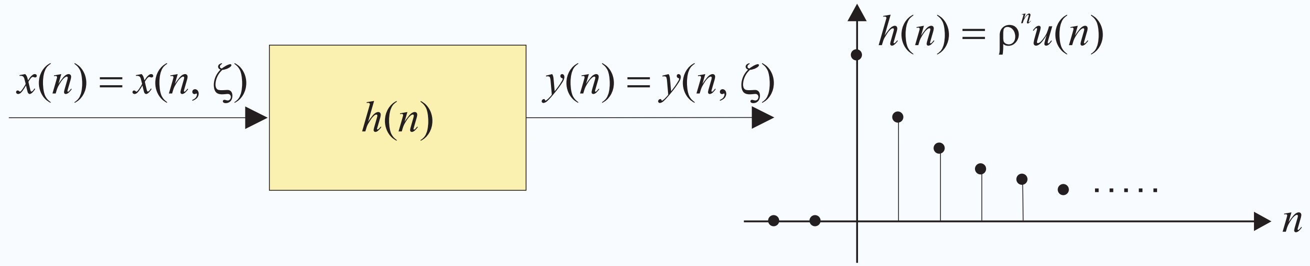
SOLUTION. Moreover, the complex spectral density is given by:

$$P_{yy}(z) = H(z)\; P_{xy}(z)$$

$$= \left(\frac{1}{1-\rho z^{-1}}\right)\left(\frac{1}{1-\rho^* z}\right)\left[\sigma_x^2 + 2\pi|\mu_x|^2\delta(z-1)\right]$$

□

Frequency-Domain Analysis of LTI systems

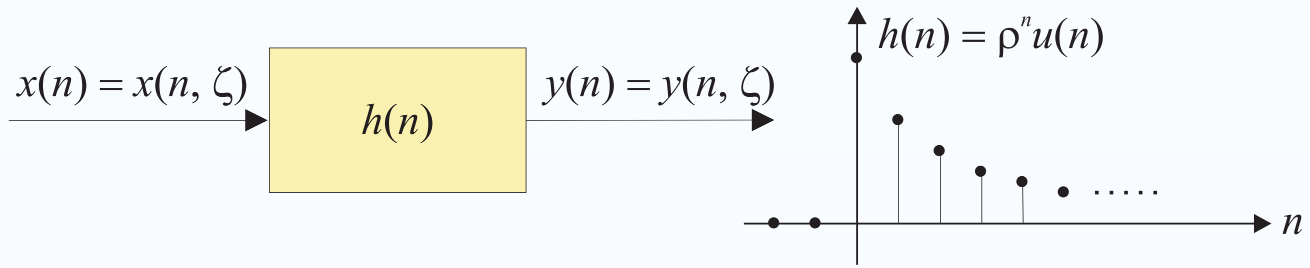


Example (Simple Example:). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2\,\delta[\ell]$. Calculate the PSD, CPSD and the complex spectral densities.

SOLUTION. Moreover, the complex spectral density is given by:

$$
\begin{aligned}
P_{yy}(z) &= H(z)\; P_{xy}(z) \\
&= \left(\frac{1}{1-\rho z^{-1}}\right)\left(\frac{1}{1-\rho^* z}\right)\left[\sigma_x^2 + 2\pi|\mu_x|^2\delta(z-1)\right] \\
&= \frac{\sigma_x^2}{1-|\rho|^2}\,\frac{1-|\rho|^2}{(1-\rho z^{-1})\,(1-\rho^* z)} + \frac{2\pi|\mu_x|^2}{|1-\rho|^2}\delta(z-1)
\end{aligned}
$$

Frequency-Domain Analysis of LTI systems



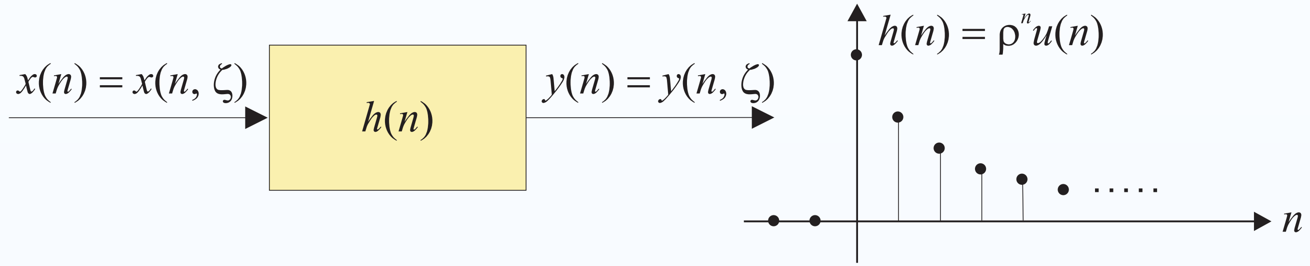
Example (Simple Example:). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2\,\delta[\ell]$. Calculate the PSD, CPSD and the complex spectral densities.

SOLUTION. The CPSD and PSD are found by setting $z = e^{j\omega}$:

$$P_{xy}(e^{j\omega}) = \frac{\sigma_x^2}{1-\rho^* e^{j\omega}} + \frac{2\pi|\mu_x|^2}{1-\rho^* e^{j\omega}}\delta(e^{j\omega}-1)$$

□

Frequency-Domain Analysis of LTI systems



Example (Simple Example:). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2\,\delta[\ell]$. Calculate the PSD, CPSD and the complex spectral densities.

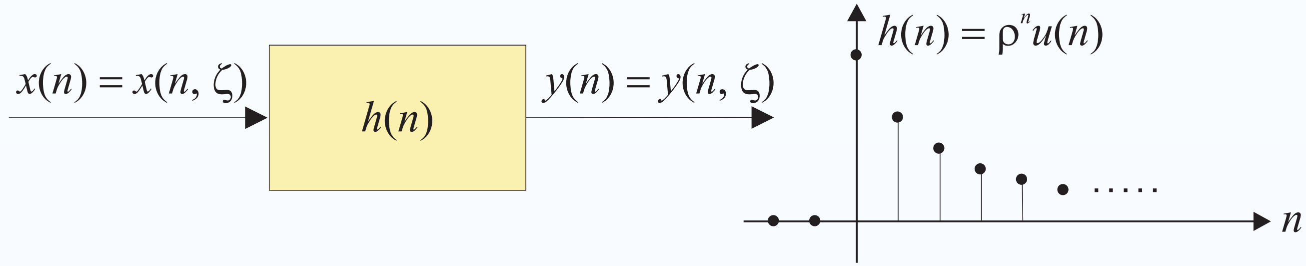
SOLUTION. The CPSD and PSD are found by setting $z = e^{j\omega}$:

$$P_{xy}(e^{j\omega}) = \frac{\sigma_x^2}{1-\rho^* e^{j\omega}} + \frac{2\pi|\mu_x|^2}{1-\rho^* e^{j\omega}}\delta(e^{j\omega}-1)$$

Moreover, the PSD is given by:

$$P_{yy}(e^{j\omega}) = \frac{\sigma_x^2}{1-|\rho|^2}\,\frac{1-|\rho|^2}{1+|\rho|^2-2|\rho|\cos(\omega-\arg\rho)} + \frac{2\pi|\mu_x|^2}{|1-\rho|^2}\delta(e^{j\omega}-1)$$

Frequency-Domain Analysis of LTI systems



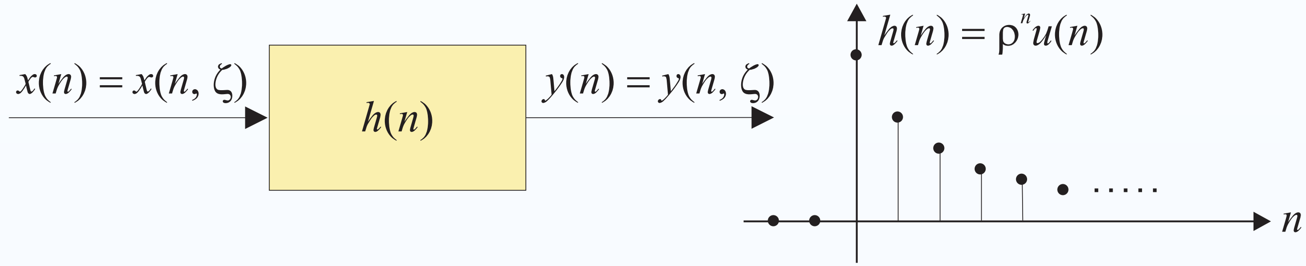
Example (Simple Example:). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2\, \delta[\ell]$. Calculate the PSD, CPSD and the complex spectral densities.

SOLUTION. Taking inverse z-transforms gives the output ACS:

$$r_{yy}[\ell] = \frac{\sigma_x^2}{1 - |\rho|^2}\, \rho^{|\ell|} + \frac{|\mu_x|^2}{|1-\rho|^2}$$

□

Frequency-Domain Analysis of LTI systems



Example (Simple Example:). The LTI system is driven by a process with mean μ_x and covariance sequence $\gamma_{xx}[\ell] = \sigma_x^2\,\delta[\ell]$. Calculate the PSD, CPSD and the complex spectral densities.

SOLUTION. Taking inverse z-transforms gives the output ACS:

$$r_{yy}[\ell] = \frac{\sigma_x^2}{1-|\rho|^2}\,\rho^{|\ell|} + \frac{|\mu_x|^2}{|1-\rho|^2} \qquad \square$$

This matches the solutions found using: the impulse response approach, or the difference equation approach.

Frequency-Domain Analysis of LTI systems

Example (Partial Fractions Example). The signal $y[n]$ is applied to the input of a system with output $s[n]$ which is characterised by:

$$s[n] = \rho\, s[n-1] + y[n] + y[n-1]$$

$\bowtie$

Frequency-Domain Analysis of LTI systems

Example (Partial Fractions Example). The signal $y[n]$ is applied to the input of a system with output $s[n]$ which is characterised by:

$$s[n] = \rho\, s[n-1] + y[n] + y[n-1]$$

- Show that the cross-power spectral density is given by:

$$P_{sy}(z) = \frac{\sigma_x^2}{1 - \rho z^{-1}} \left\{ \frac{1 + z^{-1}}{(1 - \rho\, z^{-1})\,(1 - \rho\, z)} \right\} \qquad \bowtie$$

Frequency-Domain Analysis of LTI systems

Example (Partial Fractions Example). The signal $y[n]$ is applied to the input of a system with output $s[n]$ which is characterised by:

$$s[n] = \rho\, s[n-1] + y[n] + y[n-1]$$

- Show that the cross-power spectral density is given by:

$$P_{sy}(z) = \frac{\sigma_x^2}{1 - \rho z^{-1}} \left\{ \frac{1 + z^{-1}}{(1 - \rho\, z^{-1})\,(1 - \rho\, z)} \right\} \qquad \bowtie$$

- Hence, find the cross-covariance sequence, $\gamma_{sy}[\ell]$, between the output, $s[n]$, and the input $y[n]$.

Frequency-Domain Analysis of LTI systems

Example (Partial Fractions Example). The signal $y[n]$ is applied to the input of a system with output $s[n]$ which is characterised by:

$$s[n] = \rho\, s[n-1] + y[n] + y[n-1]$$

- Show that the cross-power spectral density is given by:

$$P_{sy}(z) = \frac{\sigma_x^2}{1 - \rho z^{-1}} \left\{ \frac{1 + z^{-1}}{(1 - \rho\, z^{-1})\,(1 - \rho\, z)} \right\}$$

- Hence, find the cross-covariance sequence, $\gamma_{sy}[\ell]$, between the output, $s[n]$, and the input $y[n]$.

The following bilateral z-transform might be useful:

$$\ell\, a^{\ell}\, u[\ell] \overset{z}{\rightleftharpoons} \frac{a\, z^{-1}}{(1 - a\, z^{-1})^2}, \qquad |a| < 1$$

$\Join$

where $u[\ell] = 1$ if $\ell \geq 0$ and zero otherwise.

Frequency-Domain Analysis of LTI systems

Example (Partial Fractions Example). The signal $y[n]$ is :

$$s[n] = \rho\, s[n-1] + y[n] + y[n-1]$$

- Show that the cross-power spectral density is given by:

$$P_{sy}(z) = \frac{\sigma_x^2}{1 - \rho z^{-1}} \left\{ \frac{1 + z^{-1}}{(1 - \rho\, z^{-1})\,(1 - \rho\, z)} \right\}$$

- Hence, find $\gamma_{sy}[\ell]$.

SOLUTION. ● The cross-complex spectral density at the output:

$$P_{sy}(z) = G(z)\; P_{yy}(z) \qquad \square$$

where $G(z)$ is the transfer function of the system.

Frequency-Domain Analysis of LTI systems

Example (Partial Fractions Example). The signal $y[n]$ is :

$$s[n] = \rho\, s[n-1] + y[n] + y[n-1]$$

- Show that the cross-power spectral density is given by:

$$P_{sy}(z) = \frac{\sigma_x^2}{1 - \rho z^{-1}} \left\{ \frac{1 + z^{-1}}{(1 - \rho\, z^{-1})\,(1 - \rho\, z)} \right\}$$

- Hence, find $\gamma_{sy}[\ell]$.

SOLUTION. • By taking z-transforms:

$$G\,(z) = \frac{1 + z^{-1}}{1 - \rho z^{-1}}$$

□

Frequency-Domain Analysis of LTI systems

Example (Partial Fractions Example). The signal $y[n]$ is :

$$s[n] = \rho\, s[n-1] + y[n] + y[n-1]$$

- Show that the cross-power spectral density is given by:

$$P_{sy}(z) = \frac{\sigma_x^2}{1 - \rho z^{-1}} \left\{ \frac{1 + z^{-1}}{(1 - \rho\, z^{-1})\,(1 - \rho\, z)} \right\}$$

- Hence, find $\gamma_{sy}[\ell]$.

SOLUTION. • Using the expression for $P_{yy}(z)$:

$$P_{sy}(z) = G(z)\, P_{yy}(z) = \frac{1 + z^{-1}}{1 - \rho z^{-1}} \frac{\sigma_x^2}{(1 - \rho\, z^{-1})\,(1 - \rho\, z)} \quad \square$$

Frequency-Domain Analysis of LTI systems

Example (Partial Fractions Example). The signal $y[n]$ is :

$$s[n] = \rho\, s[n-1] + y[n] + y[n-1]$$

- Hence, find $\gamma_{sy}[\ell]$.

SOLUTION. • The term in the curly brackets can be simplified:

$$\frac{1+z^{-1}}{(1-\rho\, z^{-1})\,(1-\rho\, z)} = \frac{z+1}{(z-\rho)\,(1-\rho\, z)} = \frac{A}{z-\rho} + \frac{B}{1-\rho\, z}$$

□

Frequency-Domain Analysis of LTI systems

Example (Partial Fractions Example). The signal $y[n]$ is :

$$s[n] = \rho\, s[n-1] + y[n] + y[n-1]$$

- Hence, find $\gamma_{sy}[\ell]$.

SOLUTION. The term in the curly brackets can be simplified:

$$\frac{1+z^{-1}}{(1-\rho\, z^{-1})\,(1-\rho\, z)} = \frac{z+1}{(z-\rho)\,(1-\rho\, z)} = \frac{A}{z-\rho} + \frac{B}{1-\rho\, z}$$

Using the cover-up rule to find:

$$A\text{: } \times \text{ by } z-\rho \text{ \& set } z-\rho=0; = \frac{z+1}{(1-\rho\, z)} = A + \underbrace{(z-\rho)\,\frac{B}{1-\rho\, z}}_{=0}$$

□

Frequency-Domain Analysis of LTI systems

Example (Partial Fractions Example). The signal $y[n]$ is :

$$s[n] = \rho\, s[n-1] + y[n] + y[n-1]$$

- Hence, find $\gamma_{sy}[\ell]$.

SOLUTION.
- The term in the curly brackets can be simplified:

$$\frac{1+z^{-1}}{(1-\rho\, z^{-1})\,(1-\rho\, z)} = \frac{z+1}{(z-\rho)\,(1-\rho\, z)} = \frac{A}{z-\rho} + \frac{B}{1-\rho\, z}$$

Using the cover-up rule to find:

$$B\colon \times \text{ by } 1-\rho\, z \;\&\; \text{set } 1-\rho\, z = 0; = \frac{z+1}{(z-\rho)} = \underbrace{(1-\rho\, z)\,\frac{A}{z-\rho}}_{=0} + B$$

Frequency-Domain Analysis of LTI systems

Example (Partial Fractions Example). The signal $y[n]$ is :

$$s[n] = \rho\, s[n-1] + y[n] + y[n-1]$$

- Hence, find $\gamma_{sy}[\ell]$.

SOLUTION. ● The term in the curly brackets can be simplified:

$$\frac{1+z^{-1}}{(1-\rho\, z^{-1})\,(1-\rho\, z)} = \frac{z+1}{(z-\rho)\,(1-\rho\, z)} = \frac{A}{z-\rho} + \frac{B}{1-\rho\, z}$$

Using the cover-up rule to find:

$$A = \left.\frac{z+1}{1-\rho\, z}\right|_{z=\rho} = \frac{1+\rho}{1-\rho^2} = \frac{1}{1-\rho}$$

$$B = \left.\frac{z+1}{z-\rho}\right|_{z=\frac{1}{\rho}} = \frac{1+\rho}{1-\rho^2} = \frac{1}{1-\rho} = A$$

Frequency-Domain Analysis of LTI systems

Example (Partial Fractions Example). The signal $y[n]$ is :

$$s[n] = \rho\, s[n-1] + y[n] + y[n-1]$$

- Hence, find $\gamma_{sy}[\ell]$.

SOLUTION. • Hence, the cross-complex spectral density is:

$$P_{sy}(z) = \frac{\sigma_w^2}{1-\rho z^{-1}} \frac{1}{1-\rho} \left\{ \frac{1}{z-\rho} + \frac{1}{1-\rho\, z} \right\}$$

□

Frequency-Domain Analysis of LTI systems

Example (Partial Fractions Example). The signal $y[n]$ is :

$$s[n] = \rho\, s[n-1] + y[n] + y[n-1]$$

- Hence, find $\gamma_{sy}[\ell]$.

SOLUTION. Hence, the cross-complex spectral density is:

$$P_{sy}(z) = \frac{\sigma_w^2}{1-\rho z^{-1}} \frac{1}{1-\rho} \left\{ \frac{1}{z-\rho} + \frac{1}{1-\rho\, z} \right\}$$

$$= \frac{\sigma_w^2}{1-\rho} \left\{ \frac{1}{\rho} \frac{\rho z^{-1}}{(1-\rho\, z^{-1})^2} + \frac{1}{1-\rho^2} \frac{1-\rho^2}{(1-\rho\, z)\,(1-\rho\, z^{-1})} \right\}$$

□

Frequency-Domain Analysis of LTI systems

Example (Partial Fractions Example). The signal $y[n]$ is :

$$s[n] = \rho\, s[n-1] + y[n] + y[n-1]$$

- Hence, find $\gamma_{sy}[\ell]$.

SOLUTION. Hence, the cross-complex spectral density is:

$$P_{sy}(z) = \frac{\sigma_w^2}{1-\rho z^{-1}} \frac{1}{1-\rho} \left\{ \frac{1}{z-\rho} + \frac{1}{1-\rho\, z} \right\}$$

$$= \frac{\sigma_w^2}{1-\rho} \left\{ \frac{1}{\rho} \frac{\rho z^{-1}}{(1-\rho\, z^{-1})^2} + \frac{1}{1-\rho^2} \frac{1-\rho^2}{(1-\rho\, z)\,(1-\rho\, z^{-1})} \right\}$$

Hence, taking inverse-z-transforms gives the cross-covariance:

Frequency-Domain Analysis of LTI systems

Example (Partial Fractions Example). The signal $y[n]$ is :

$$s[n] = \rho\, s[n-1] + y[n] + y[n-1]$$

- Hence, find $\gamma_{sy}[\ell]$.

SOLUTION.
- Hence, taking inverse-z-transforms gives the cross-covariance:

$$\gamma_{sy}[\ell] = \frac{\sigma_w^2}{1-\rho}\left\{\frac{\ell}{\rho}\,\rho^{\ell}\,u[\ell] + \frac{1}{1-\rho^2}\,\rho^{|\ell|}\right\}$$

□

Frequency-Domain Analysis of LTI systems

Example (Partial Fractions Example). The signal $y[n]$ is :

$$s[n] = \rho\, s[n-1] + y[n] + y[n-1]$$

- Hence, find $\gamma_{sy}[\ell]$.

SOLUTION.
- Hence, taking inverse-z-transforms gives the cross-covariance:

$$\gamma_{sy}[\ell] = \frac{\sigma_w^2}{1-\rho} \left\{ \frac{\ell}{\rho} \rho^\ell\, u[\ell] + \frac{1}{1-\rho^2} \rho^{|\ell|} \right\}$$

  - To find the cross-correlation requires the addition of the mean components as before.
  - To find the output auto-correlation requires substantially more work, and this is left as an exercise to the reader! □

Frequency-Domain Analysis of LTI systems

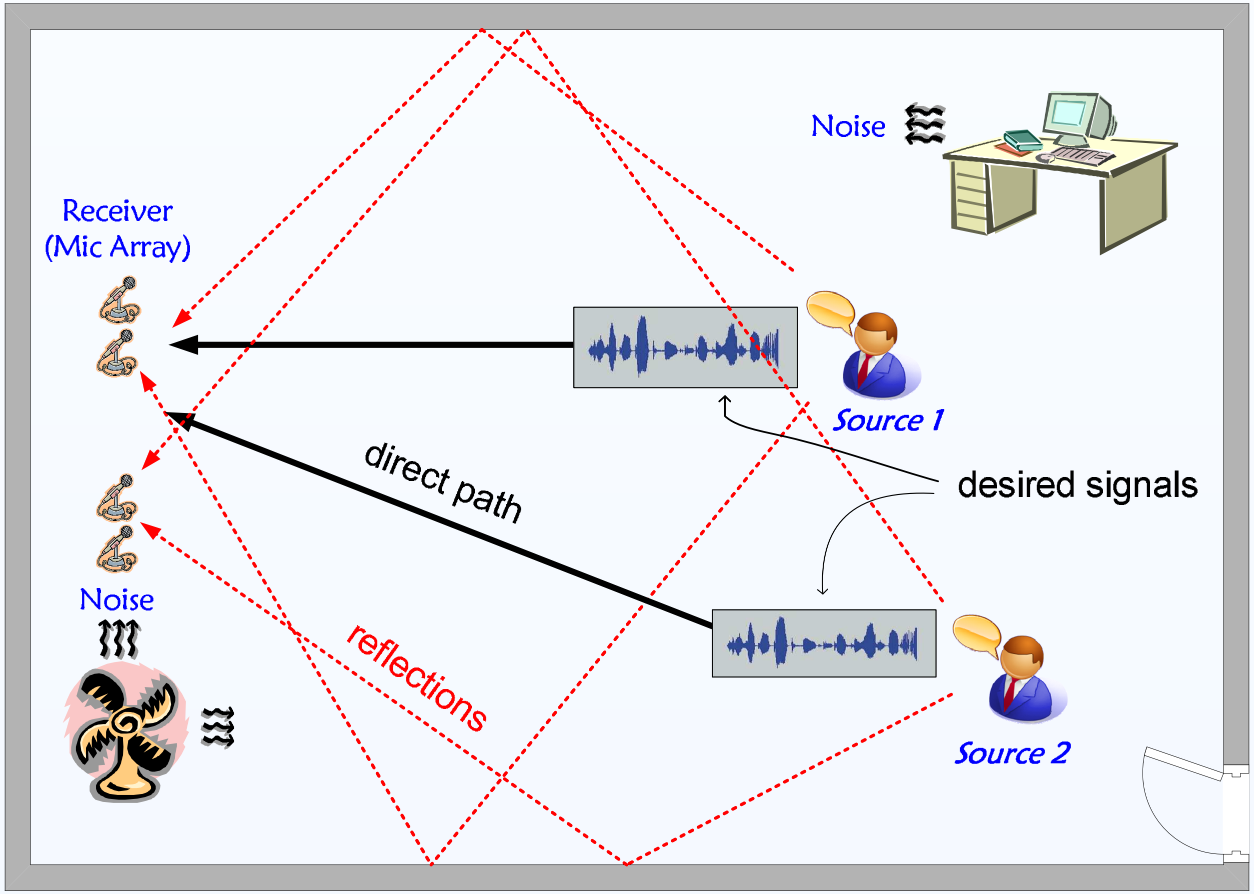
– End-of-Topic 61: **Frequency-domain analysis of input-output statistics** –

Any Questions?

Advanced Topics

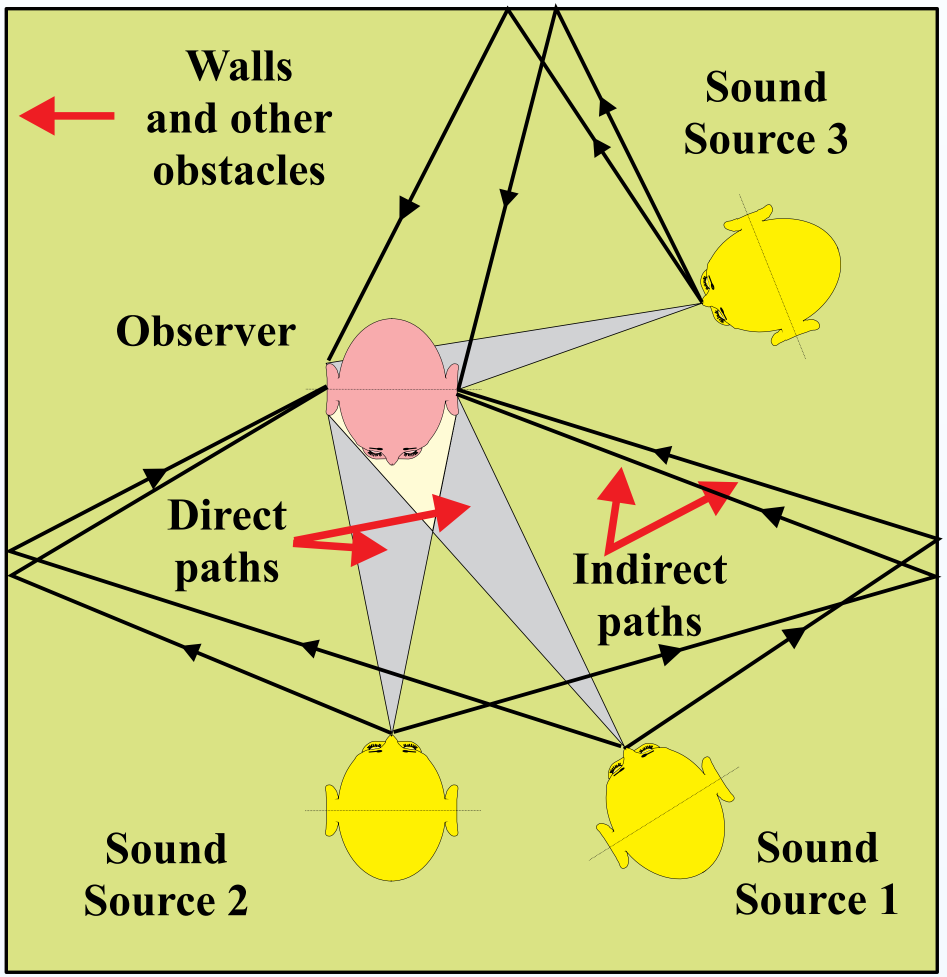
Lecture Slideset 1
Passive Target Localisation

Introduction



Source localisation and BSS.

Introduction



Humans turn their head in the direction of interest in order to reduce interference from other directions; *joint detection, localisation, and enhancement*.

Introduction

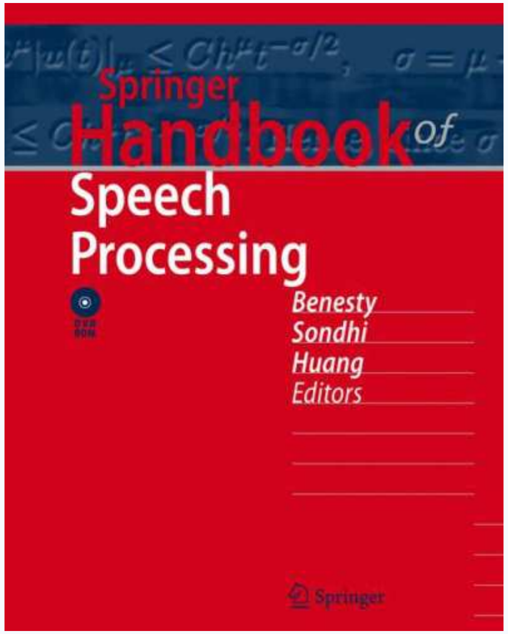
- This research tutorial is intended to cover a wide range of aspects which link acoustic source localisation (ASL) and blind source separation (BSS).
- This tutorial is being continually updated, and feedback is welcomed. The documents published on the USB stick may differ to the slides presented on the day.
- The latest version of this document can be found online and downloaded at:

  `http://mod-udrc.org/events/2016-summer-school`
- Thanks to Xionghu Zhong and Ashley Hughes for borrowing some of their diagrams from their dissertations.

Structure of the Tutorial

- Recommended Texts
- Conceptual link between ASL and BSS.
- Geometry of source localisation.
- Spherical and hyperboloidal localisation.
- Estimating TDOAs.
- Steered beamformer response function.
- Multiple target localisation using BSS.
- Conclusions.

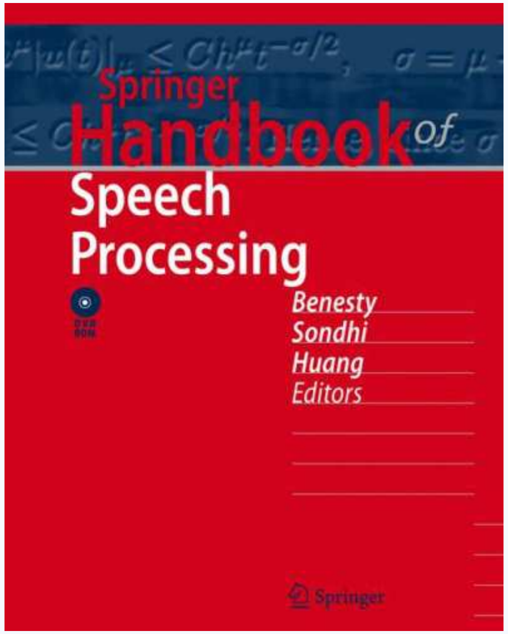
Recommended Texts

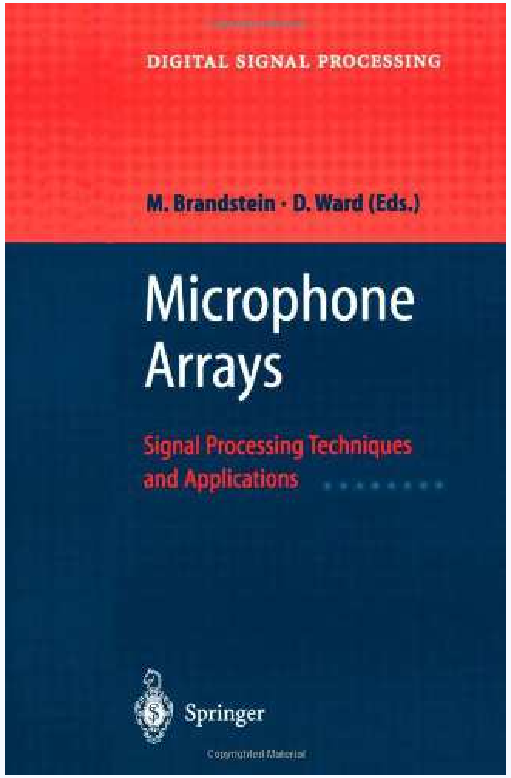


Recommended book chapters and the references therein.

- Huang Y., J. Benesty, and J. Chen, "Time Delay Estimation and Source Localization," in *Springer Handbook of Speech Processing* by J. Benesty, M. M. Sondhi, and Y. Huang, pp. 1043–1063, , Springer, 2008.

Recommended Texts

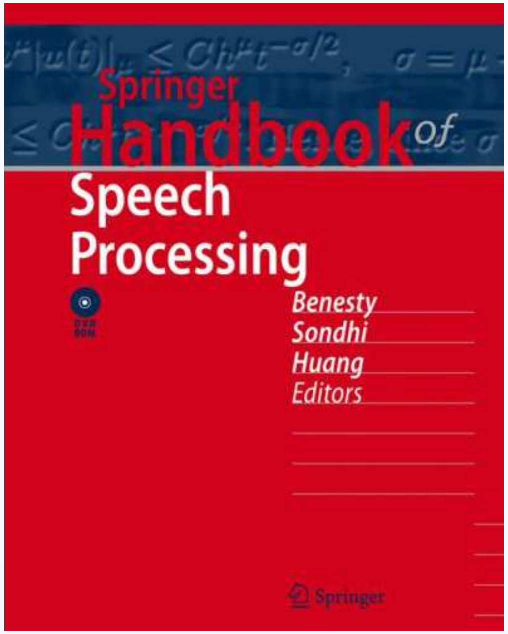


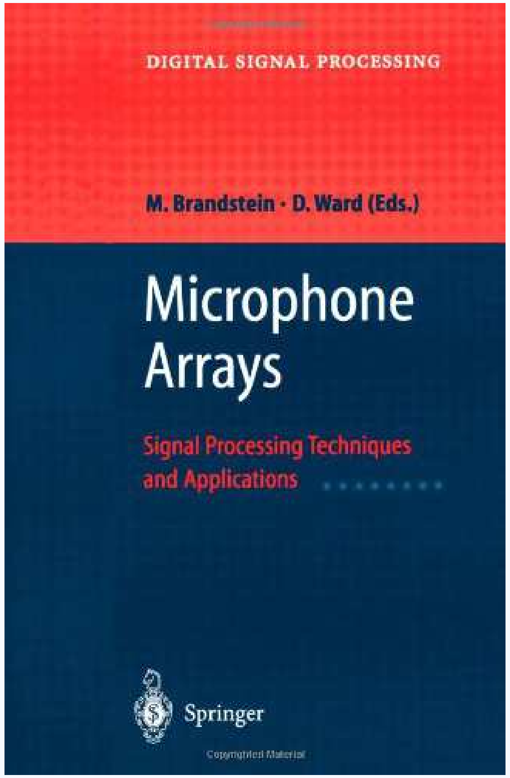


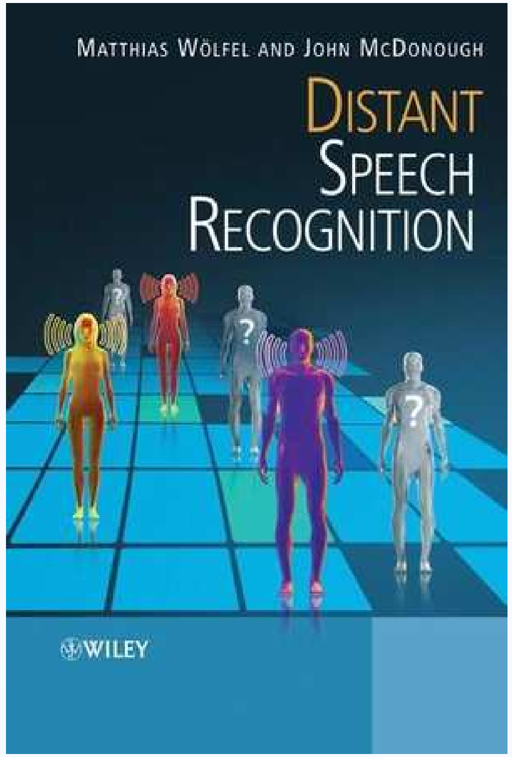
Recommended book chapters and the references therein.

- Chapter 8: DiBiase J. H., H. F. Silverman, and M. S. Brandstein, "Robust Localization in Reverberant Rooms," in *Microphone Arrays* by M. Brandstein and D. Ward, pp. 157–180, , Springer Berlin Heidelberg, 2001.

Recommended Texts







Recommended book chapters and the references therein.

- Chapter 10 of Wolfel M. and J. McDonough, *Distant Speech Recognition*, Wiley, 2009.

  IDENTIFIERS – *Hardback*, ISBN13: 978-0-470-51704-8

Recommended Texts

Some recent PhD thesis on the topic include:

- Zhong X., *"Bayesian framework for multiple acoustic source tracking,"* Ph.D. thesis, University of Edinburgh, 2010.
- Pertila P., *"Acoustic Source Localization in a Room Environment and at Moderate Distances,"* Ph.D. thesis, Tampere University of Technology, 2009.
- Fallon M., *"Acoustic Source Tracking using Sequential Monte Carlo,"* Ph.D. thesis, University of Cambridge, 2008.

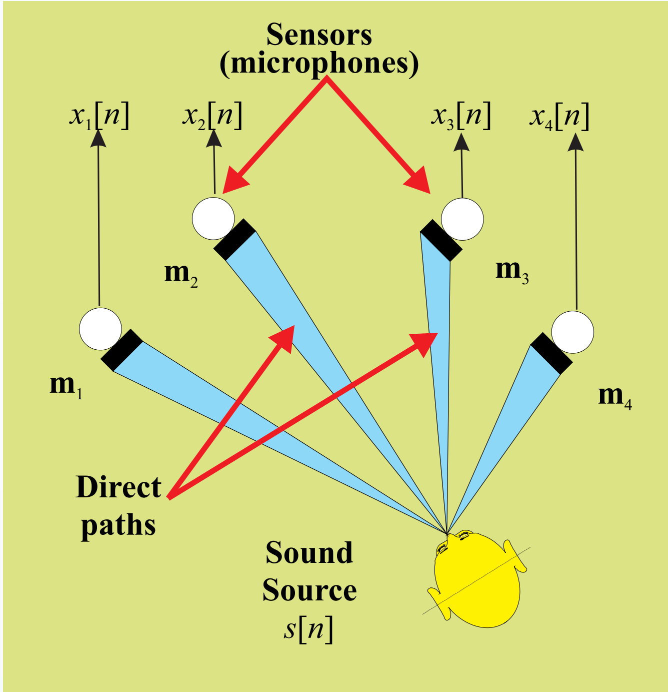
Why Source Localisation?

A number of blind source separation (BSS) techniques rely on knowledge of the desired source position:

1. Look-direction in beamforming techniques.
2. Camera steering for audio-visual BSS (including Robot Audition).
3. Parametric modelling of the mixing matrix.

Equally, a number of multi-target acoustic source localisation (ASL) techniques rely on BSS.

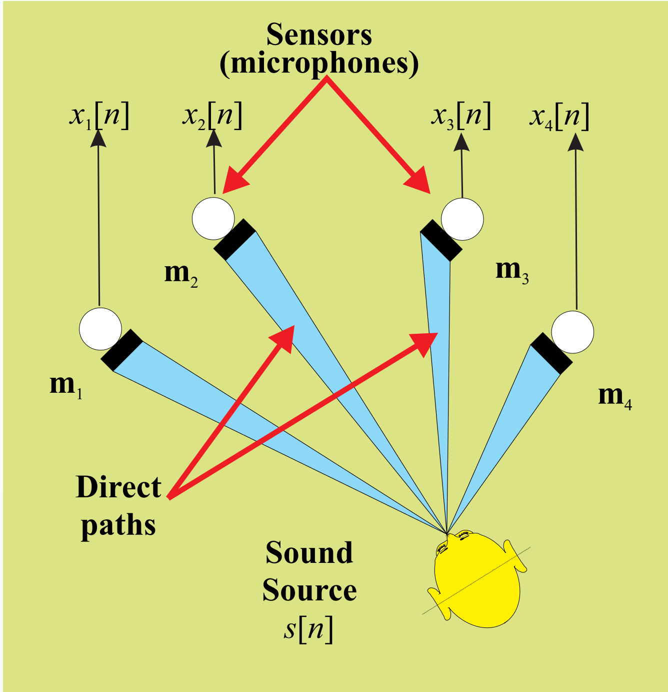
ASL Methodology



Ideal free-field model.

- Most ASL techniques rely on the fact that an impinging wavefront reaches one sensor before it reaches another.

ASL Methodology



Ideal free-field model.

- Most ASL techniques rely on the fact that an impinging wavefront reaches one sensor before it reaches another.
- Most ASL algorithms are designed assuming there is no reverberation present, the *free-field assumption*.

ASL Methodology

An uniform linear array (ULA) of microphones.

- Typically, this acoustic sensor is a microphone; will primarily consider *omni-directional pressure sensors*, and rely on the TDOA between the signals at different microphones.

ASL Methodology

An ULA of microphones.

- Typically, this acoustic sensor is a microphone; will primarily consider *omni-directional pressure sensors*, and rely on the TDOA between the signals at different microphones.
- Other measurement types include:
  - range difference measurements;
  - interaural level difference;
  - joint TDOA and vision techniques.

ASL Methodology

- Another sensor modality might include acoustic vector sensors (AVSs) which measure both air pressure and air velocity. Useful for applications such as sniper localisation.

An acoustic vector sensor.

Source Localization Strategies

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

1. those based on maximising the SRP of a beamformer;

   - location estimate derived directly from a filtered, weighted, and sum version of the signal data.

Source Localization Strategies

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

1. those based on maximising the SRP of a beamformer;
   - location estimate derived directly from a filtered, weighted, and sum version of the signal data.
2. techniques adopting high-resolution spectral estimation concepts (see Stephan Weiss's talk);
   - any localisation scheme relying upon an application of the signal correlation matrix.

Source Localization Strategies

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

1. those based on maximising the SRP of a beamformer;
   - location estimate derived directly from a filtered, weighted, and sum version of the signal data.
2. techniques adopting high-resolution spectral estimation concepts (see Stephan Weiss's talk);
   - any localisation scheme relying upon an application of the signal correlation matrix.
3. approaches employing TDOA information.
   - source locations calculated from a set of TDOA estimates measured across various combinations of microphones.

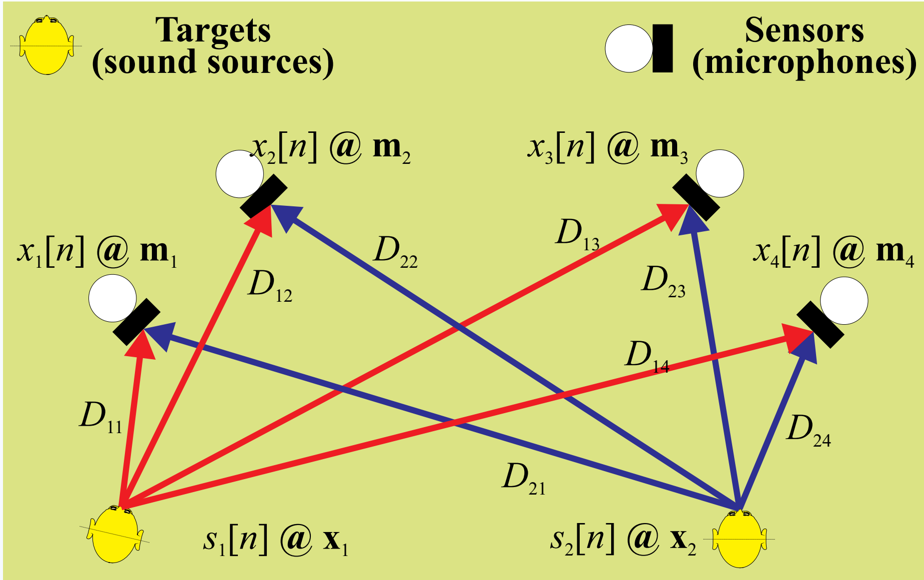
Source Localization Strategies

Spectral-estimation approaches See Stephan Weiss's talk :-)

TDOA-based estimators Computationally cheap, but suffers in the presence of noise and reverberation.

SBF approaches Computationally intensive, superior performance to TDOA-based methods. However, possible to dramatically reduce computational load.

Geometric Layout

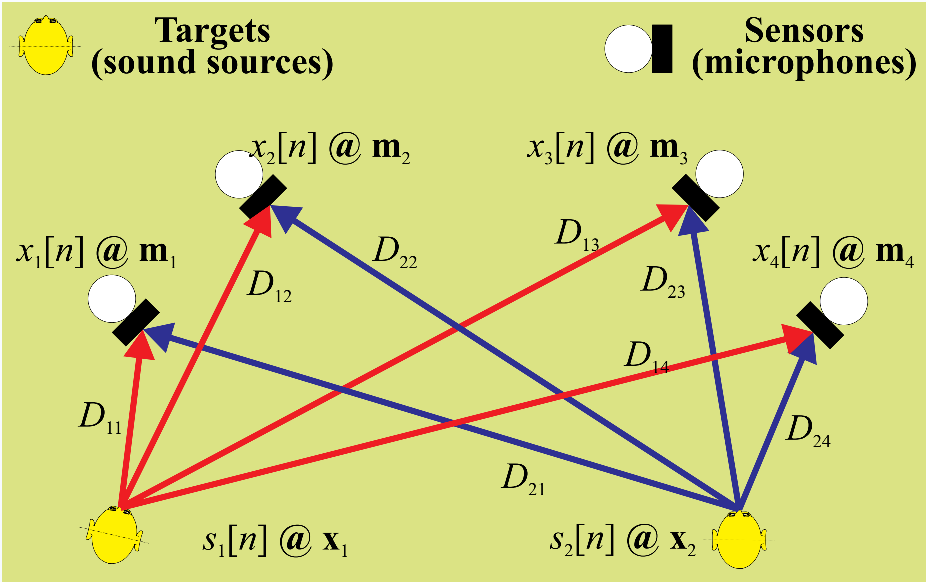


Geometry assuming a free-field model.

Suppose there is a:

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

Geometric Layout



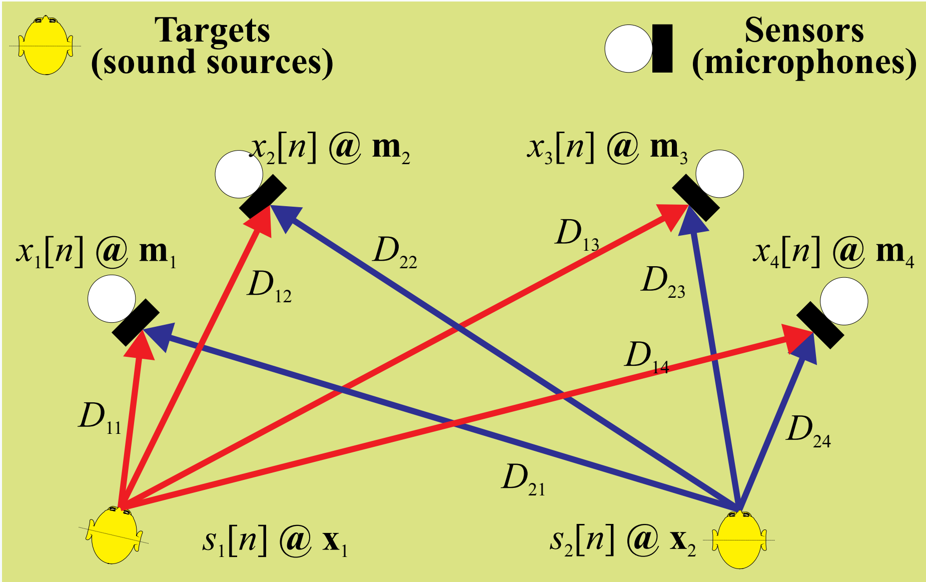
Geometry assuming a free-field model.

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

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

where c is the speed of sound, which is approximately 344 m/s.

Geometric Layout



Geometry assuming a free-field model.

The distance from the target at $\mathbf{x}_k$ to the sensor located at $\mathbf{m}_i$ will be defined by D_{ik}, and is called the range.

$$T_{ij}(\mathbf{x}_k) = \frac{1}{c}(D_{ik} - D_{jk})$$

Ideal Free-field Model

- In an anechoic free-field acoustic environment, the signal from source k, denoted by $s_k(t)$, propagates to the i-th sensor at time t according to the expression:

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

where $b_{ik}(t)$ denotes additive noise. Note that, in the frequency domain, this expression is given by:

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

- The additive noise source is assumed to be uncorrelated with the source signal, as well as the noise signals at the other microphones.

Ideal Free-field Model

- In an anechoic free-field acoustic environment, the signal from source k, denoted by $s_k(t)$, propagates to the i-th sensor at time t according to the expression:

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

  where $b_{ik}(t)$ denotes additive noise. Note that, in the frequency domain, this expression is given by:

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

- The additive noise source is assumed to be uncorrelated with the source signal, as well as the noise signals at the other microphones.
- The TDOA between the i-th and j-th microphone is given by:

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

TDOA and Hyperboloids

It is important to be aware of the geometrical properties that arise from the TDOA relationship

$$T(\mathbf{m}_i, \mathbf{m}_j, \mathbf{x}_k) = \frac{|\mathbf{x}_k - \mathbf{m}_i| - |\mathbf{x}_k - \mathbf{m}_j|}{c}$$

TDOA and Hyperboloids

It is important to be aware of the geometrical properties that arise from the TDOA relationship

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

- This defines one half of a hyperboloid of two sheets, centered on the midpoint of the microphones, $\mathbf{v}_{ij} = \frac{\mathbf{m}_i + \mathbf{m}_j}{2}$.

$$\left(\mathbf{x}_k - \mathbf{v}_{ij}\right)^T \mathbf{V}_{ij} \left(\mathbf{x}_k - \mathbf{v}_{ij}\right) = 1$$

TDOA and Hyperboloids

It is important to be aware of the geometrical properties that arise from the TDOA relationship

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

- This defines one half of a hyperboloid of two sheets, centered on the midpoint of the microphones, $\mathbf{v}_{ij}=\frac{\mathbf{m}_i+\mathbf{m}_j}{2}$.

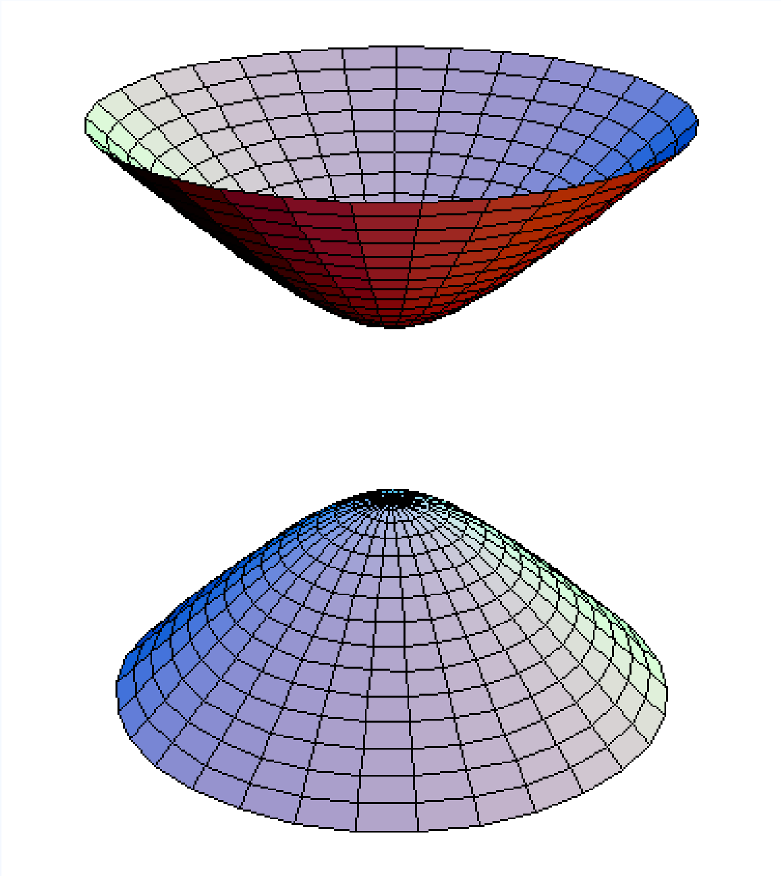
$$\left(\mathbf{x}_k-\mathbf{v}_{ij}\right)^T\mathbf{V}_{ij}\left(\mathbf{x}_k-\mathbf{v}_{ij}\right)=1$$

- For source with a large source-range to microphone-separation ratio, the hyperboloid may be well-approximated by a cone with a constant direction angle relative to the axis of symmetry.

$$\phi_{ij}=\cos^{-1}\left(\frac{c\,T\left(\mathbf{m}_i,\,\mathbf{m}_j,\,\mathbf{x}_k\right)}{\left|\mathbf{m}_i-\mathbf{m}_j\right|}\right)$$

TDOA and Hyperboloids

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

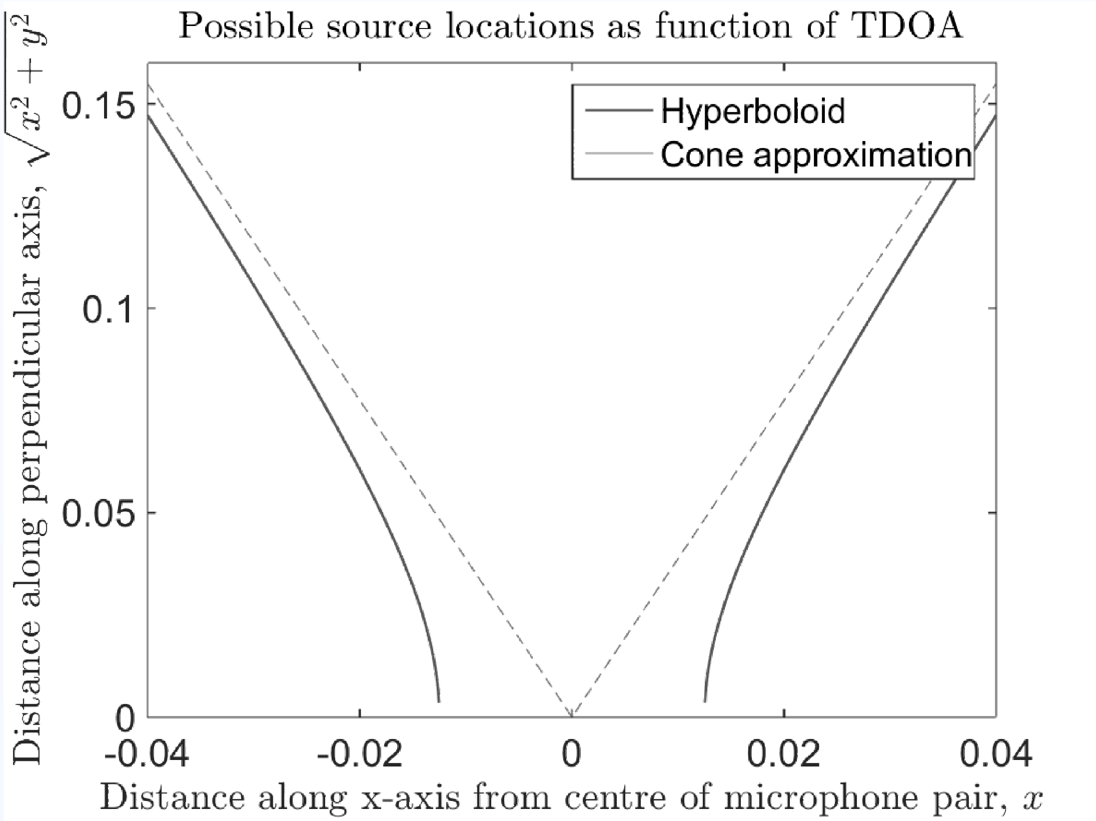


Hyperboloid of two sheets

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = -1$$

TDOA and Hyperboloids

$$T(\mathbf{m}_i, \mathbf{m}_j, \mathbf{x}_k) = \frac{|\mathbf{x}_k - \mathbf{m}_i| - |\mathbf{x}_k - \mathbf{m}_j|}{c}$$



Hyperboloid, for a microphone separation of $d = 0.1$, and a time-delay of $\tau_{ij} = \frac{d}{4c}$.

Indirect TDOA-based Methods

This is typically a two-step procedure in which:

- Typically, TDOAs are extracted using the GCC function, or an AED algorithm.

Indirect TDOA-based Methods

This is typically a two-step procedure in which:

- Typically, TDOAs are extracted using the GCC function, or an AED algorithm.
- A hypothesised spatial position of the target can be used to predict the expected TDOAs (or corresponding range) at the microphone.

Indirect TDOA-based Methods

This is typically a two-step procedure in which:

- Typically, TDOAs are extracted using the GCC function, or an AED algorithm.
- A hypothesised spatial position of the target can be used to predict the expected TDOAs (or corresponding range) at the microphone.
- The error between the measured and hypothesised TDOAs is then minimised.

Indirect TDOA-based Methods

This is typically a two-step procedure in which:

- Typically, TDOAs are extracted using the GCC function, or an AED algorithm.
- A hypothesised spatial position of the target can be used to predict the expected TDOAs (or corresponding range) at the microphone.
- 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 ASL methods.

Indirect TDOA-based Methods

This is typically a two-step procedure in which:

- Typically, TDOAs are extracted using the GCC function, or an AED algorithm.
- A hypothesised spatial position of the target can be used to predict the expected TDOAs (or corresponding range) at the microphone.
- 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 ASL methods.
- An alternative way of viewing these solutions is to consider what **spatial positions** of the target could lead to the estimated TDOA.

Spherical Least Squares Error Function

- Suppose the first microphone is located at the origin of the coordinate system, such that $\mathbf{m_0} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^T$.

- The range from target k to sensor i can be expressed as :

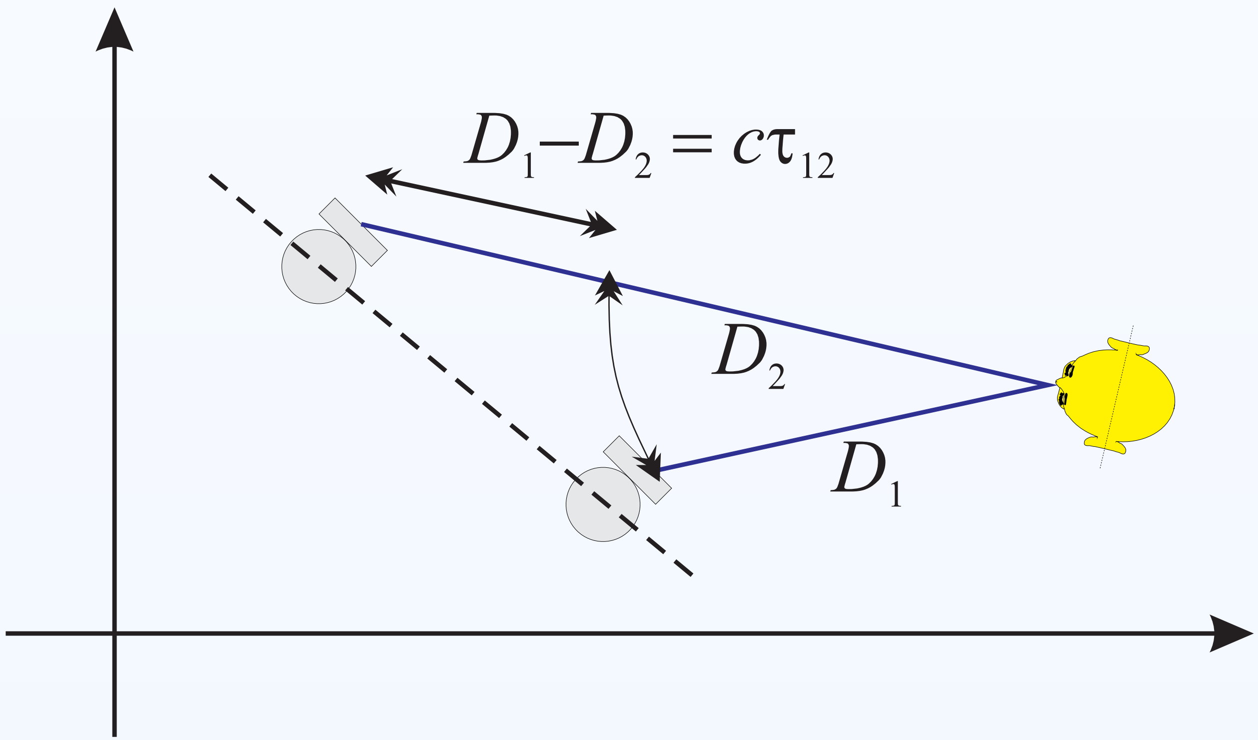
$$\begin{aligned} D_{ik} &= D_{0k} + D_{ik} - D_{0k} \\ &= R_s + c\,T_{i0}\left(\mathbf{x}_k\right) \end{aligned}$$

where $R_{sk} = |\mathbf{x}_k|$ is the range to the first microphone which is at the origin.

Spherical Least Squares Error Function

- In practice, the observations are the TDOAs and, given R_{sk}, these ranges can be considered the **measurement ranges**.

  Of course, knowing R_{sk} is half the solution, but it is just one unknown at this stage.



Range and TDOA relationship.

Spherical Least Squares Error Function

- The source-sensor geometry states that the target lies on a sphere centered on the corresponding sensor. Hence,

$$\begin{aligned} D_{ik}^2 &= |\mathbf{x}_k - \mathbf{m}_i|^2 \\ &= \mathbf{x}_k^T \mathbf{x}_k - 2\mathbf{m}_i^T \mathbf{x}_k + \mathbf{m}_i^T \mathbf{m}_i \\ &= R_s^2 - 2\mathbf{m}_i^T \mathbf{x}_k + R_i^2 \end{aligned}$$

$R_i = |\mathbf{m}_i|$ is the distance of the i-th microphone to the origin.

Spherical Least Squares Error Function

- The source-sensor geometry states that the target lies on a sphere centered on the corresponding sensor. Hence,

$$
\begin{aligned}
D_{ik}^2 &= |\mathbf{x}_k - \mathbf{m}_i|^2 \\
&= \mathbf{x}_k^T \mathbf{x}_k - 2\mathbf{m}_i^T \mathbf{x}_k + \mathbf{m}_i^T \mathbf{m}_i \\
&= R_s^2 - 2\mathbf{m}_i^T \mathbf{x}_k + R_i^2
\end{aligned}
$$

- Define the **spherical error function** as:

$$
\epsilon_{ik} \triangleq \frac{1}{2}\left(\hat{D}_{ik}^2 - D_{ik}^2\right)
$$

Spherical Least Squares Error Function

- The source-sensor geometry states that the target lies on a sphere centered on the corresponding sensor. Hence,

$$\begin{aligned} D_{ik}^2 &= \left|\mathbf{x}_k - \mathbf{m}_i\right|^2 \\ &= \mathbf{x}_k^T\,\mathbf{x}_k - 2\mathbf{m}_i^T\,\mathbf{x}_k + \mathbf{m}_i^T\,\mathbf{m}_i \\ &= R_s^2 - 2\mathbf{m}_i^T\,\mathbf{x}_k + R_i^2 \end{aligned}$$

- Define the **spherical error function** as:

$$\begin{aligned} \epsilon_{ik} &\triangleq \frac{1}{2}\left(\hat{D}_{ik}^2 - D_{ik}^2\right) \\ &= \frac{1}{2}\left\{\left(R_s + c\,\hat{T}_{i0}\right)^2 - \left(R_s^2 - 2\mathbf{m}_i^T\,\mathbf{x}_k + R_i^2\right)\right\} \end{aligned}$$

Spherical Least Squares Error Function

- The source-sensor geometry states that the target lies on a sphere centered on the corresponding sensor. Hence,

$$
\begin{aligned}
D_{ik}^2 &= \left|\mathbf{x}_k - \mathbf{m}_i\right|^2 \\
&= \mathbf{x}_k^T \mathbf{x}_k - 2\mathbf{m}_i^T \mathbf{x}_k + \mathbf{m}_i^T \mathbf{m}_i \\
&= R_s^2 - 2\mathbf{m}_i^T \mathbf{x}_k + R_i^2
\end{aligned}
$$

- Define the **spherical error function** as:

$$
\begin{aligned}
\epsilon_{ik} &\triangleq \frac{1}{2}\left(\hat{D}_{ik}^2 - D_{ik}^2\right) \\
&= \frac{1}{2}\left\{\left(R_s + c\,\hat{T}_{i0}\right)^2 - \left(R_s^2 - 2\mathbf{m}_i^T \mathbf{x}_k + R_i^2\right)\right\} \\
&= \mathbf{m}_i^T \mathbf{x_k} + c\,R_s\,\hat{T}_{i0} + \frac{1}{2}\left(c^2 \hat{T}_{i0}^2 - R_i^2\right)
\end{aligned}
$$

Spherical Least Squares Error Function

- Concatenating the error functions for each microphone gives the expression:

$$\boldsymbol{\epsilon}_{ik} = \mathbf{A}\,\mathbf{x}_k - \underbrace{(\mathbf{b}_k - R_{sk}\mathbf{d}_k)}_{\mathbf{v}_k}$$

$$\equiv \underbrace{\begin{bmatrix} \mathbf{A} & \mathbf{d}_k \end{bmatrix}}_{\mathbf{S}_k} \underbrace{\begin{bmatrix} \mathbf{x}_k \\ R_{sk} \end{bmatrix}}_{\boldsymbol{\theta}_k} - \mathbf{b}_k$$

where

$$\mathbf{A} = \begin{bmatrix} \mathbf{m}_0^T \\ \vdots \\ \mathbf{m}_{N-1}^T \end{bmatrix}, \mathbf{d} = c \begin{bmatrix} \hat{T}_{00} \\ \vdots \\ \hat{T}_{(N-1)0} \end{bmatrix}, \quad \mathbf{b}_k = \frac{1}{2} \begin{bmatrix} c^2 \hat{T}_{00}^2 - R_0^2 \\ \vdots \\ c^2 \hat{T}_{(N-1)0}^2 - R_{N-1}^2 \end{bmatrix}$$

Spherical Least Squares Error Function

- The LSE can then be obtained by using $J = \boldsymbol{\epsilon}_i^T \boldsymbol{\epsilon}_i$:

$$J(\mathbf{x}_k) = (\mathbf{A}\mathbf{x}_k - (\mathbf{b}_k - R_{sk}\,\mathbf{d}_k))^T (\mathbf{A}\mathbf{x}_k - (\mathbf{b}_k - R_{sk}\,\mathbf{d}_k))$$

$$J(\mathbf{x}_k,\, \boldsymbol{\theta}_k) = (\mathbf{S}_k\boldsymbol{\theta}_k - \mathbf{b}_k)^T (\mathbf{S}_k\boldsymbol{\theta}_k - \mathbf{b}_k)$$

Spherical Least Squares Error Function

- The LSE can then be obtained by using $J = \boldsymbol{\epsilon}_i^T \boldsymbol{\epsilon}_i$:

$$J(\mathbf{x}_k) = (\mathbf{A}\mathbf{x}_k - (\mathbf{b}_k - R_{sk}\,\mathbf{d}_k))^T (\mathbf{A}\mathbf{x}_k - (\mathbf{b}_k - R_{sk}\,\mathbf{d}_k))$$

$$J(\mathbf{x}_k,\, \boldsymbol{\theta}_k) = (\mathbf{S}_k\boldsymbol{\theta}_k - \mathbf{b}_k)^T (\mathbf{S}_k\boldsymbol{\theta}_k - \mathbf{b}_k)$$

- Note that as $R_{sk} = |\mathbf{x}_k|$, these parameters aren't independent. Therefore, the problem can either be formulated as:
  - a nonlinear least-squares problem in $\mathbf{x}_k$;
  - a linear minimisation subject to quadratic constraints:

$$\hat{\boldsymbol{\theta}}_k = \arg\min_{\boldsymbol{\theta}_k} (\mathbf{S}_k\boldsymbol{\theta}_k - \mathbf{b}_k)^T (\mathbf{S}_k\boldsymbol{\theta}_k - \mathbf{b}_k)$$

subject to the constraint

$$\boldsymbol{\theta}_k\,\Delta\,\boldsymbol{\theta}_k = 0 \qquad \text{where} \qquad \Delta = \text{diag}\,[1,\,1,\,1,\,-1]$$

Spherical Least Squares Error Function

Two-step Spherical LSE Approaches

To avoid solving either a nonlinear or a constrained least-squares problem, it is possible to solve the problem in two steps, namely:

1. solving a LLS problem in $\mathbf{x}_k$ *assuming* the range to the target, R_{sk}, is known;

2. and then solving for R_{sk} given an estimate of $\mathbf{x}_k$ i. t. o. R_{sk}.

Two-step Spherical LSE Approaches

To avoid solving either a nonlinear or a constrained least-squares problem, it is possible to solve the problem in two steps, namely:

1. solving a LLS problem in $\mathbf{x}_k$ *assuming* the range to the target, R_{sk}, is known;

2. and then solving for R_{sk} given an estimate of $\mathbf{x}_k$ i. t. o. R_{sk}.

- Assuming an estimate of R_{sk} this can be solved as

$$\hat{\mathbf{x}}_k = \mathbf{A}^\dagger \mathbf{v}_k = \mathbf{A}^\dagger \left(\mathbf{b}_k - \hat{R}_{sk} \mathbf{d}_k \right) \quad \text{where} \quad \mathbf{A}^\dagger = \left[\mathbf{A}^T \mathbf{A} \right]^{-1} \mathbf{A}^T$$

Note that $\mathbf{A}^\dagger$ is the pseudo-inverse of $\mathbf{A}$.

Spherical Intersection Estimator

This method uses the physical constraint that the range R_{sk} is the Euclidean distance to the target.

- Writing $\hat{R}_{sk}^2 = \hat{\mathbf{x}}_k^T \, \hat{\mathbf{x}}_k$, it follows that:

$$\hat{R}_{sk}^2 = \left(\mathbf{b}_k - \hat{R}_{sk}\mathbf{d}_k\right)^T \mathbf{A}^{\dagger\,T}\mathbf{A}^{\dagger} \left(\mathbf{b}_k - \hat{R}_{sk}\mathbf{d}_k\right)$$

Spherical Intersection Estimator

This method uses the physical constraint that the range R_{sk} is the Euclidean distance to the target.

- Writing $\hat{R}_{sk}^2 = \hat{\mathbf{x}}_k^T \, \hat{\mathbf{x}}_k$, it follows that:

$$\hat{R}_{sk}^2 = \left(\mathbf{b}_k - \hat{R}_{sk}\mathbf{d}_k\right)^T \mathbf{A}^{\dagger\,T}\mathbf{A}^{\dagger} \left(\mathbf{b}_k - \hat{R}_{sk}\mathbf{d}_k\right)$$

which can be written as the quadratic:

$$a\,\hat{R}_{sk}^2 + b\,\hat{R}_{sk} + c = 0$$

Spherical Intersection Estimator

This method uses the physical constraint that the range R_{sk} is the Euclidean distance to the target.

- Writing $\hat{R}_{sk}^2 = \hat{\mathbf{x}}_k^T \, \hat{\mathbf{x}}_k$, it follows that:

$$\hat{R}_{sk}^2 = \left(\mathbf{b}_k - \hat{R}_{sk}\mathbf{d}_k\right)^T \mathbf{A}^{\dagger\,T}\mathbf{A}^{\dagger} \left(\mathbf{b}_k - \hat{R}_{sk}\mathbf{d}_k\right)$$

which can be written as the quadratic:

$$a\,\hat{R}_{sk}^2 + b\,\hat{R}_{sk} + c = 0$$

- The unique, real, positive root is taken as the spherical intersection (SX) estimator of the source range. Hence, the estimator will fail when:

1. there is no real, positive root, or:

2. if there are two positive real roots.

Spherical Interpolation Estimator

The spherical interpolation (SI) estimator again uses the spherical least squares error (LSE) function, but this time the range R_{sk} is estimated in the least-squares sense.

Consider again the **spherical error function**:

$$\boldsymbol{\epsilon}_{ik} = \mathbf{A}\mathbf{x}_k - (\mathbf{b}_k - R_{sk}\,\mathbf{d}_k)$$

Spherical Interpolation Estimator

The SI estimator again uses the spherical LSE function, but this time the range R_{sk} is estimated in the least-squares sense.

Consider again the **spherical error function**:

$$\boldsymbol{\epsilon}_{ik} = \mathbf{A}\mathbf{x}_k - (\mathbf{b}_k - R_{sk}\,\mathbf{d}_k)$$

Substituting the LSE gives:

$$\boldsymbol{\epsilon}_{ik} = \mathbf{A}\left[\mathbf{A}^T\mathbf{A}\right]^{-1}\mathbf{A}^T\left(\mathbf{b}_k - \hat{R}_{sk}\mathbf{d}_k\right) - (\mathbf{b}_k - R_{sk}\,\mathbf{d}_k)$$

Spherical Interpolation Estimator

The SI estimator again uses the spherical LSE function, but this time the range R_{sk} is estimated in the least-squares sense.

Consider again the **spherical error function**:

$$\boldsymbol{\epsilon}_{ik} = \mathbf{A}\mathbf{x}_k - (\mathbf{b}_k - R_{sk}\,\mathbf{d}_k)$$

Substituting the LSE gives:

$$\boldsymbol{\epsilon}_{ik} = \mathbf{A}\left[\mathbf{A}^T\mathbf{A}\right]^{-1}\mathbf{A}^T\left(\mathbf{b}_k - \hat{R}_{sk}\mathbf{d}_k\right) - (\mathbf{b}_k - R_{sk}\,\mathbf{d}_k)$$

Defining the projection matrix as $\mathbf{P}_\mathbf{A} = \mathbf{I}_N - \mathbf{A}\left[\mathbf{A}^T\mathbf{A}\right]^{-1}\mathbf{A}^T$,

$$\boldsymbol{\epsilon}_{ik} = R_{sk}\,\mathbf{P}_\mathbf{A}\mathbf{d}_k - \mathbf{P}_\mathbf{A}\mathbf{b}_k$$

Spherical Interpolation Estimator

The SI estimator again uses the spherical LSE function, but this time the range R_{sk} is estimated in the least-squares sense.

Consider again the **spherical error function**:

$$\boldsymbol{\epsilon}_{ik} = \mathbf{A}\mathbf{x}_k - (\mathbf{b}_k - R_{sk}\,\mathbf{d}_k)$$

Defining the projection matrix as $\mathbf{P}_\mathbf{A} = \mathbf{I}_N - \mathbf{A}\left[\mathbf{A}^T\mathbf{A}\right]^{-1}\mathbf{A}^T$,

$$\boldsymbol{\epsilon}_{ik} = R_{sk}\,\mathbf{P}_\mathbf{A}\mathbf{d}_k - \mathbf{P}_\mathbf{A}\mathbf{b}_k$$

Minimising the LSE using the normal equations gives:

$$R_{sk} = \frac{\mathbf{d}_k^T\mathbf{P}_\mathbf{A}\mathbf{b}_k}{\mathbf{d}_k^T\,\mathbf{P}_\mathbf{A}\mathbf{d}_k}$$

Spherical Interpolation Estimator

The SI estimator again uses the spherical LSE function, but this time the range R_{sk} is estimated in the least-squares sense.

Consider again the **spherical error function**:

$$\epsilon_{ik} = \mathbf{A}\mathbf{x}_k - (\mathbf{b}_k - R_{sk}\,\mathbf{d}_k)$$

Substituting back into the LSE for the target position gives the final estimator:

$$\hat{\mathbf{x}}_k = \mathbf{A}^\dagger \left(\mathbf{I}_N - \mathbf{d}_k \frac{\mathbf{d}_k^T \mathbf{P}_\mathbf{A}}{\mathbf{d}_k^T \mathbf{P}_\mathbf{A} \mathbf{d}_k} \right) \mathbf{b}_k$$

This approach is said to perform better, but is computationally slightly more complex than the SX estimator.

Other Approaches

There are several other approaches to minimising the spherical LSE function .

- In particular, the **linear-correction** LSE solves the constrained minimization problem using Lagrange multipliers in a two stage process.

- For further information, see: Huang Y., J. Benesty, and J. Chen, "Time Delay Estimation and Source Localization," in *Springer Handbook of Speech Processing* by J. Benesty, M. M. Sondhi, and Y. Huang, pp. 1043–1063, , Springer, 2008.

Hyperbolic Least Squares Error Function

- If a TDOA is estimated between two microphones i and j, then the error between this and modelled TDOA is:

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

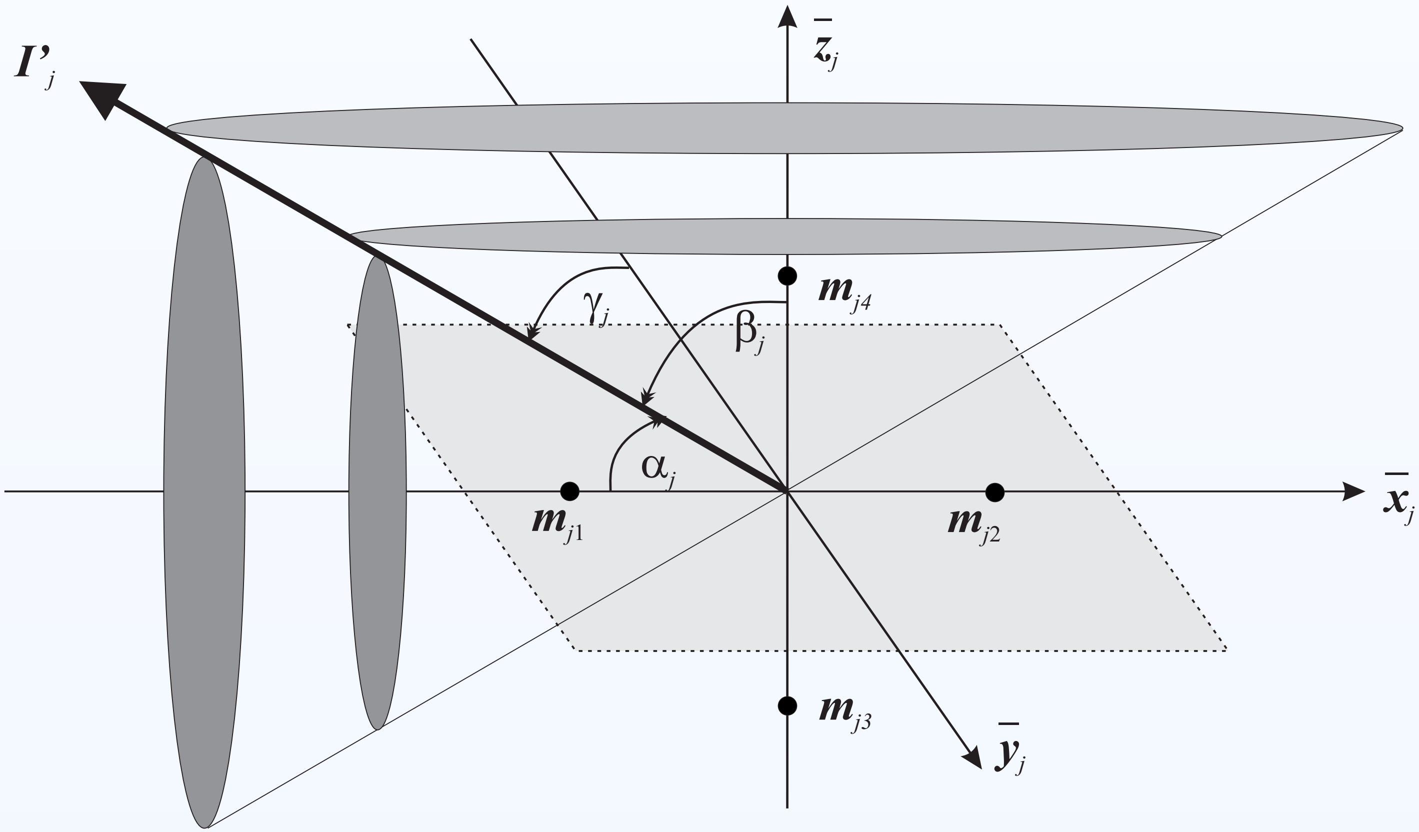
- The total error as a function of target position

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

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

Linear Intersection Method

The linear intersection (LI) algorithm works by utilising a *sensor quadruple* with a common midpoint, which allows a bearing line to be deduced from the intersection of two cones.

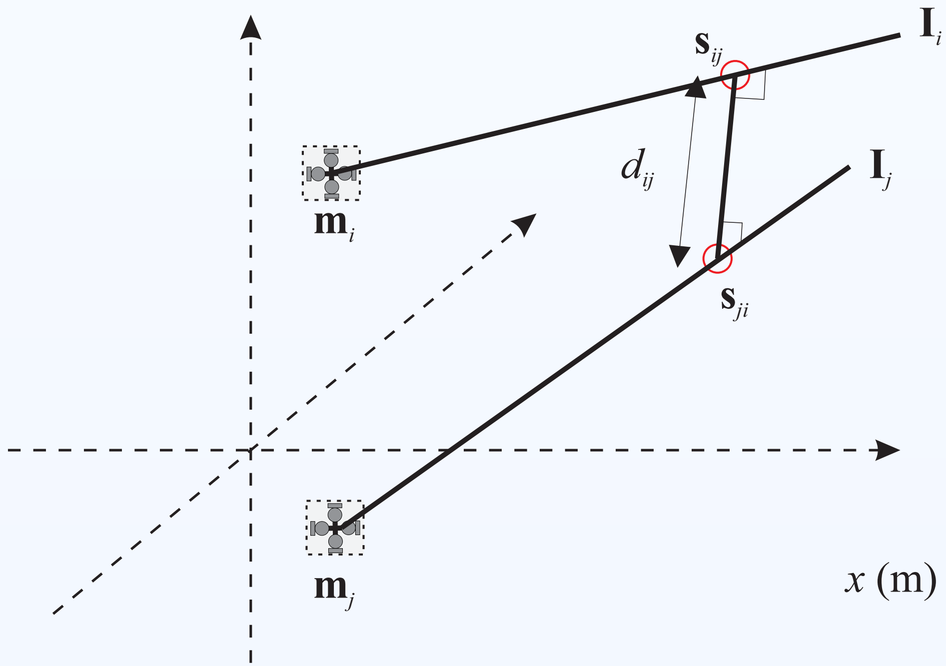


Quadruple sensor arrangement and local Cartesian coordinate system.

Linear Intersection Method

- Given the bearing lines, it is possible to calculate the points $\mathbf{s}_{ij}$ and $\mathbf{s}_{ji}$ on two bearing lines which give the closest intersection. This is basic gemoentry.

- The trick is to note that given these points $\mathbf{s}_{ij}$ and $\mathbf{s}_{ji}$, the theoretical TDOA, $T(\mathbf{m}_{1i}, \mathbf{m}_{2i}, \mathbf{s}_{ij})$, can be compared with the observed TDOA.



Calculating the points of closest intersection.

TDOA estimation methods

Two key methods for TDOA estimation are using the GCC function and the AED algorithm.

GCC algorithm most popular approach assuming an ideal free-field movel

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

TDOA estimation methods

Two key methods for TDOA estimation are using the GCC function and the AED algorithm.

GCC algorithm most popular approach assuming an ideal free-field movel

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

However, GCC-based methods

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

TDOA estimation methods

Two key methods for TDOA estimation are using the GCC function and the AED algorithm.

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

- adopts a reverberant rather than free-field model;
- computationally more expensive than GCC;
- can fail when there are common-zeros in the room impulse response (RIR).

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

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

- The cross-correlation function is given by

$$\begin{aligned} r_{x_i\, x_j}[\ell] &= \mathcal{F}^{-1}\left(\Phi\left(e^{j\omega T_s}\right) P_{x_1 x_2}\left(e^{j\omega T_s}\right)\right) \\ &= \int_{-\frac{\pi}{T_s}}^{\frac{\pi}{T_s}} \Phi\left(e^{j\omega T_s}\right) P_{x_1 x_2}\left(e^{j\omega T_s}\right) e^{j\ell\omega T}\, d\omega \end{aligned}$$

where the CPSD is given by

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

CPSD for Free-Field Model

For the free-field model , it follows that for $i \neq j$:

$$\begin{aligned} P_{x_i x_j}(\omega) &= \mathbb{E}\left[X_j(\omega)\, X_j(\omega)\right] \\ &= \mathbb{E}\left[\left(\alpha_{ik}\, S_k(\omega)\; e^{-j\omega\,\tau_{ik}} + B_{ik}(\omega)\right)\left(\alpha_{jk}\, S_k(\omega)\; e^{-j\omega\,\tau_{kk}} + B_{jk}(\omega)\right)\right] \\ &= \alpha_{ik}\alpha_{jk} e^{-j\omega\, T(\mathbf{m}_i,\, \mathbf{m}_j,\, \mathbf{x}_k)} \mathbb{E}\left[\left|S_k(\omega)\right|^2\right] \end{aligned}$$

where $\mathbb{E}\left[B_{ik}(\omega)\, B_{jk}(\omega)\right] = 0$ and $\mathbb{E}\left[B_{ik}(\omega)\, S_k(\omega)\right] = 0$.

CPSD for Free-Field Model

For the free-field model , it follows that for $i \neq j$:

$$P_{x_i x_j}(\omega) = \mathbb{E}\left[X_j(\omega)\, X_j(\omega)\right]$$
$$= \mathbb{E}\left[\left(\alpha_{ik}\, S_k(\omega)\; e^{-j\omega\,\tau_{ik}} + B_{ik}(\omega)\right)\left(\alpha_{jk}\, S_k(\omega)\; e^{-j\omega\,\tau_{kk}} + B_{jk}(\omega)\right)\right]$$
$$= \alpha_{ik}\alpha_{jk} e^{-j\omega\, T(\mathbf{m}_i,\, \mathbf{m}_j,\, \mathbf{x}_k)} \mathbb{E}\left[\left|S_k(\omega)\right|^2\right]$$

where $\mathbb{E}\left[B_{ik}(\omega)\, B_{jk}(\omega)\right] = 0$ and $\mathbb{E}\left[B_{ik}(\omega)\, S_k(\omega)\right] = 0$.

- In particular, note that it follows:

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

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.

GCC Processors

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

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

GCC Processors

The PHAT-GCC approach can be written as:

$$
\begin{aligned}
r_{x_i x_j}[\ell] &= \int_{-\frac{\pi}{T_s}}^{\frac{\pi}{T_s}} \Phi\left(e^{j\omega T_s}\right) P_{x_1 x_2}\left(e^{j\omega T_s}\right) \, e^{j\ell\omega T} \, d\omega \\
&= \int_{-\frac{\pi}{T_s}}^{\frac{\pi}{T_s}} \frac{1}{\left|P_{x_1 x_2}\left(e^{j\omega T_s}\right)\right|} \left|P_{x_1 x_2}\left(e^{j\omega T_s}\right)\right| e^{j\angle P_{x_1 x_2}\left(e^{j\omega T_s}\right)} \, e^{j\ell\omega T} \, d\omega \\
&= \int_{-\frac{\pi}{T_s}}^{\frac{\pi}{T_s}} e^{j\left(\ell\omega T + \angle P_{x_1 x_2}\left(e^{j\omega T_s}\right)\right)} \, d\omega \\
&= \delta\left(\ell\, T_s + \angle P_{x_1 x_2}\left(e^{j\omega T_s}\right)\right) \\
&= \delta(\ell\, T_s - T\left(\mathbf{m}_i,\, \mathbf{m}_j,\, \mathbf{x}_k\right))
\end{aligned}
$$

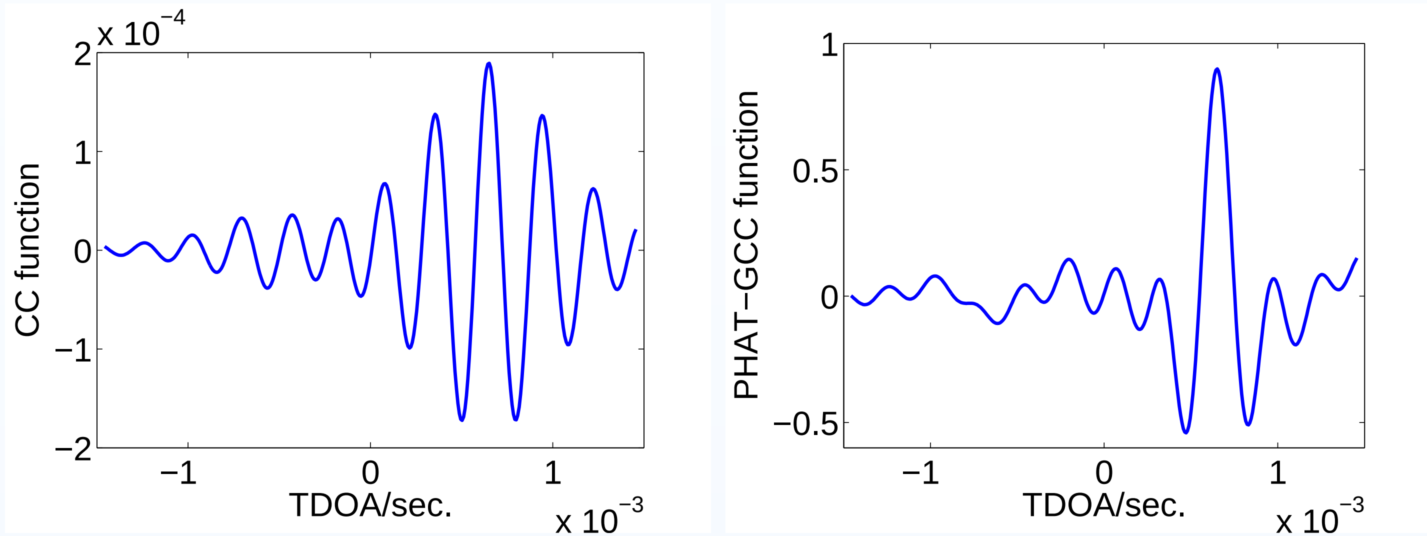
GCC Processors

The PHAT-GCC approach can be written as:

$$
\begin{aligned}
r_{x_i\,x_j}[\ell] &= \int_{-\frac{\pi}{T_s}}^{\frac{\pi}{T_s}} \Phi\left(e^{j\omega T_s}\right) P_{x_1 x_2}\left(e^{j\omega T_s}\right)\, e^{j\ell\omega T}\, d\omega \\
&= \int_{-\frac{\pi}{T_s}}^{\frac{\pi}{T_s}} \frac{1}{\left|P_{x_1 x_2}\left(e^{j\omega T_s}\right)\right|} \left|P_{x_1 x_2}\left(e^{j\omega T_s}\right)\right| e^{j\angle P_{x_1 x_2}\left(e^{j\omega T_s}\right)}\, e^{j\ell\omega T}\, d\omega \\
&= \int_{-\frac{\pi}{T_s}}^{\frac{\pi}{T_s}} e^{j\left(\ell\omega T + \angle P_{x_1 x_2}\left(e^{j\omega T_s}\right)\right)}\, d\omega \\
&= \delta\left(\ell\, T_s + \angle P_{x_1 x_2}\left(e^{j\omega T_s}\right)\right) \\
&= \delta(\ell\, T_s - T\left(\mathbf{m}_i,\, \mathbf{m}_j,\, \mathbf{x}_k\right))
\end{aligned}
$$

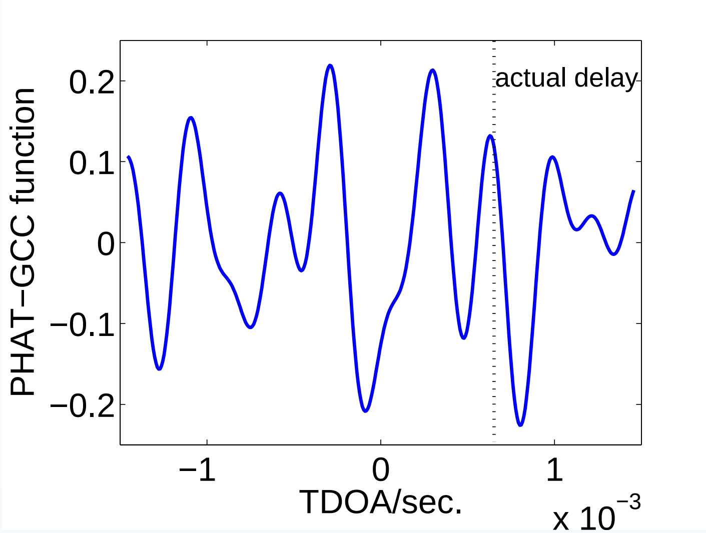
- In the absence of reverberation, the GCC-PHAT algorithm gives an impulse at a lag given by the TDOA divided by the sampling period.

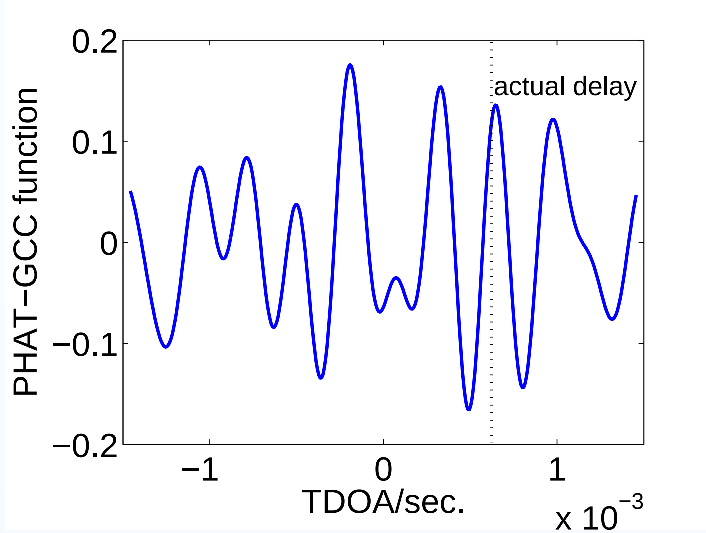
GCC Processors



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

GCC Processors





The effect of reverberation and noise on the GCC-PHAT can lead to poor TDOA estimates.

Adaptive Eigenvalue Decomposition

The AED algorithm actually amounts to a **blind channel identification** problem, which then seeks to identify the channel coefficients corresponding to the direct path elements.

Adaptive Eigenvalue Decomposition

The AED algorithm actually amounts to a **blind channel identification** problem, which then seeks to identify the channel coefficients corresponding to the direct path elements.

- Suppose that the acoustic impulse response (AIR) between source k and i is given by $h_{ik}[n]$ such that

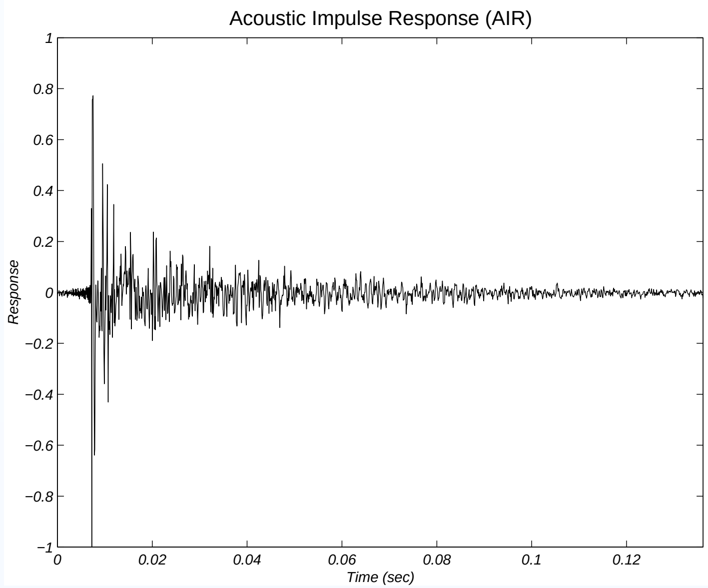
$$x_{ik}[n] = \sum_{m=-\infty}^{\infty} h_{ik}[n-m]\, s_k[m] + b_{ik}[n]$$

then the TDOA between microphones i and j is:

$$\tau_{ijk} = \left\{ \arg\max_{\ell} |h_{ik}[\ell]| \right\} - \left\{ \arg\max_{\ell} |h_{jk}[\ell]| \right\}$$

This assumes a minimum-phase system, but can easily be made robust to a non-minimum-phase system.

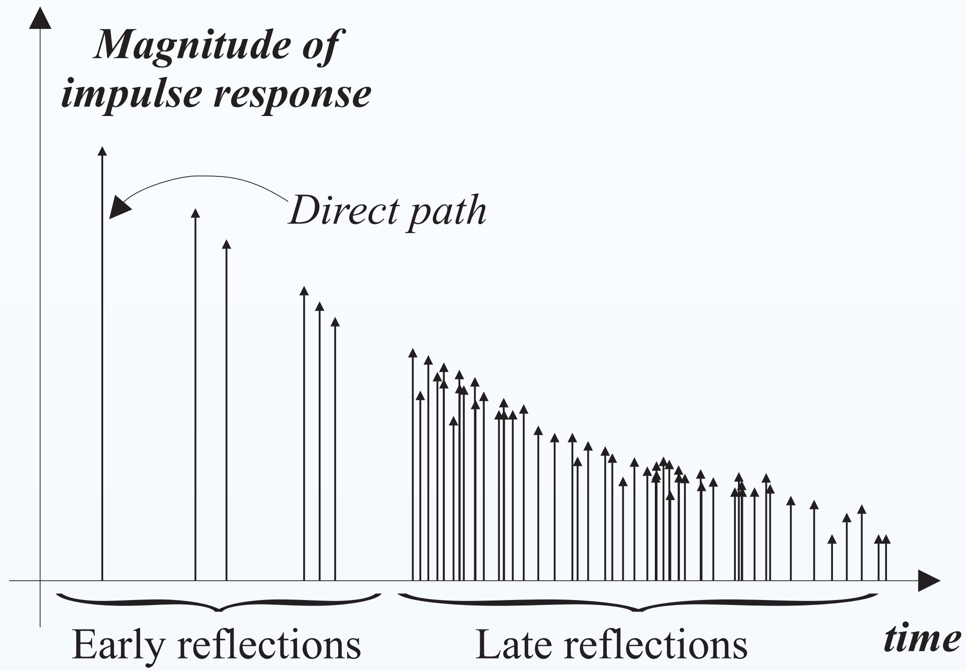
Adaptive Eigenvalue Decomposition



A typical room acoustic impulse response.

- Reverberation plays a major role in ASL and BSS.
- Consider reverberation as the sum total of all sound reflections arriving at a certain point in a room after room has been excited by impulse.

Adaptive Eigenvalue Decomposition

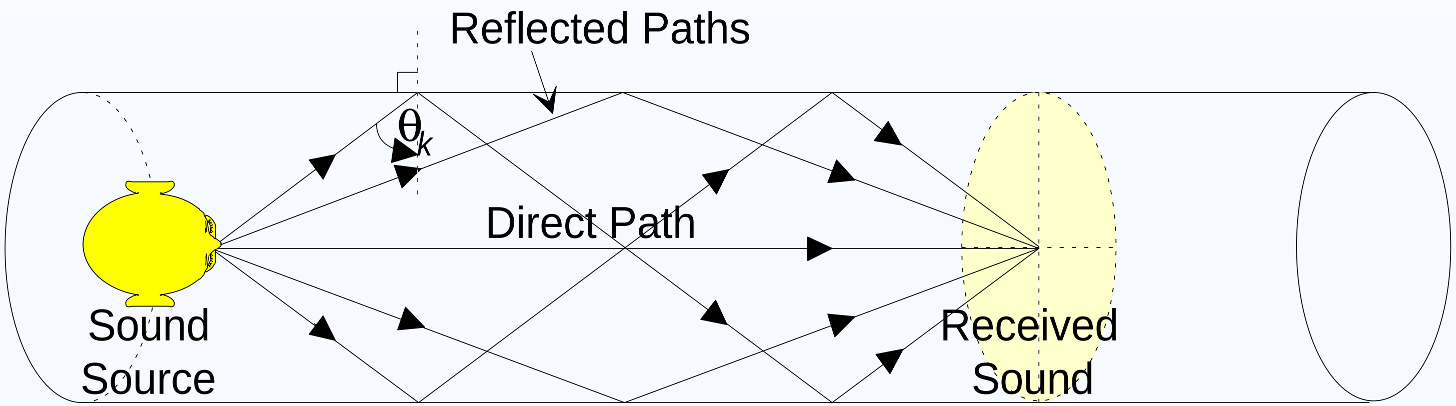


Early and late reflections in an AIR.

Trivia: Perceive early reflections to reinforce direct sound, and can help with speech intelligibility. It can be easier to hold a conversation in a closed room than outdoors

Adaptive Eigenvalue Decomposition

- Room transfer functions are often nonminimum-phase since there is more energy in the reverberant component of the RIR than in the component corresponding to direct path.



Demonstrating nonminimum-phase properties

- Therefore AED will need to consider multiple peaks in the estimated AIR.

Direct Localisation Methods

- 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.

Steered Response Power Function

The 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 $\hat{\tau}_{pk} = |\hat{\mathbf{x}} - \mathbf{m}_p|$:

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

Steered Response Power Function

The 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 $\hat{\tau}_{pk} = |\hat{\mathbf{x}} - \mathbf{m}_p|$:

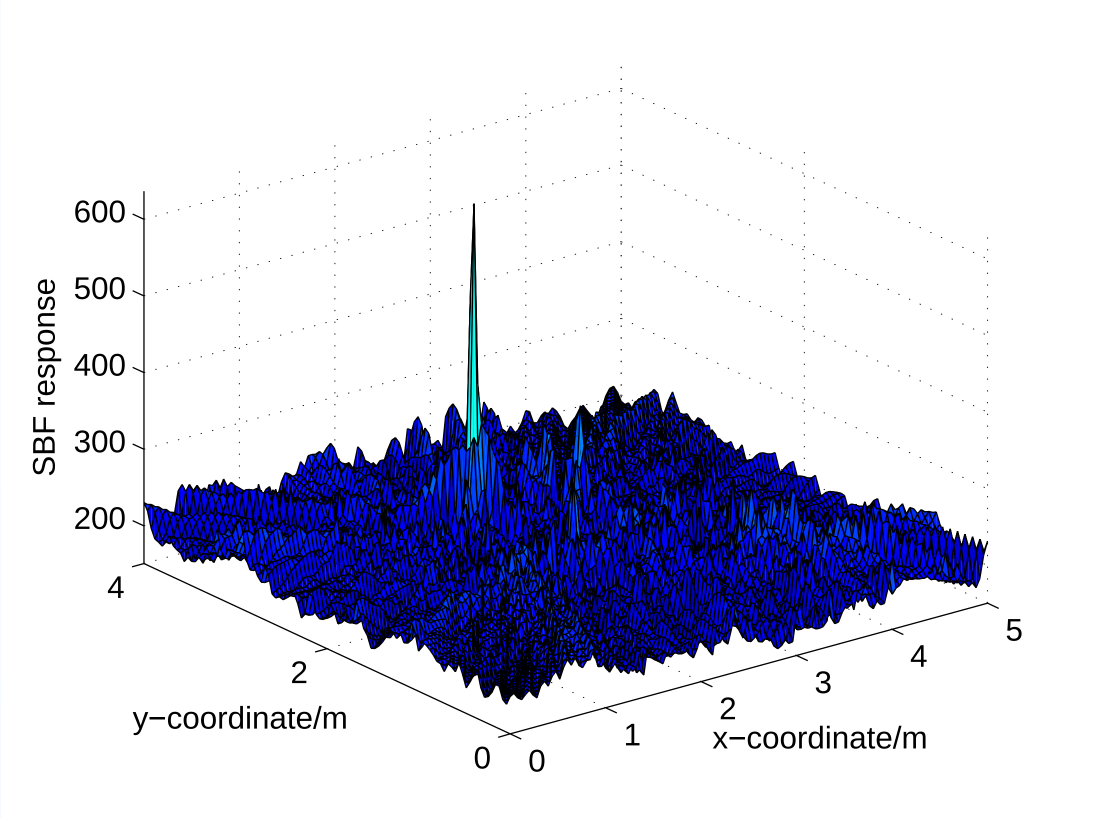
$$S(\hat{\mathbf{x}}) = \int_\Omega \left| \sum_{p=1}^{N} W_p\left(e^{j\omega T_s}\right) X_p\left(e^{j\omega T_s}\right) e^{j\omega\,\hat{\tau}_{pk}} \right|^2 d\omega$$

Taking expectations, $\Phi_{pq}\left(e^{j\omega T_s}\right) = W_p\left(e^{j\omega T_s}\right) W_q^*\left(e^{j\omega T_s}\right)$

$$\mathbb{E}\left[S(\hat{\mathbf{x}})\right] = \sum_{p=1}^{N}\sum_{q=1}^{N} \int_\Omega \Phi_{pq}\left(e^{j\omega T_s}\right) P_{x_p x_q}\left(e^{j\omega T_s}\right) e^{j\omega \hat{\tau}_{pqk}}\, d\omega$$

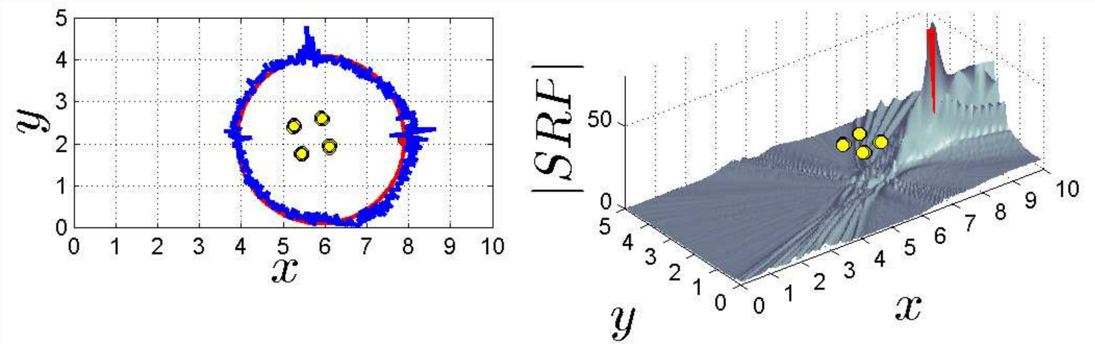
$$= \sum_{p=1}^{N}\sum_{q=1}^{N} r_{x_i\,x_j}[\hat{\tau}_{pqk}] \equiv \sum_{p=1}^{N}\sum_{q=1}^{N} r_{x_i\,x_j}\left[\frac{|\mathbf{x}_k - \mathbf{m}_i| - |\mathbf{x}_k - \mathbf{m}_j|}{c}\right]$$

Steered Response Power Function



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

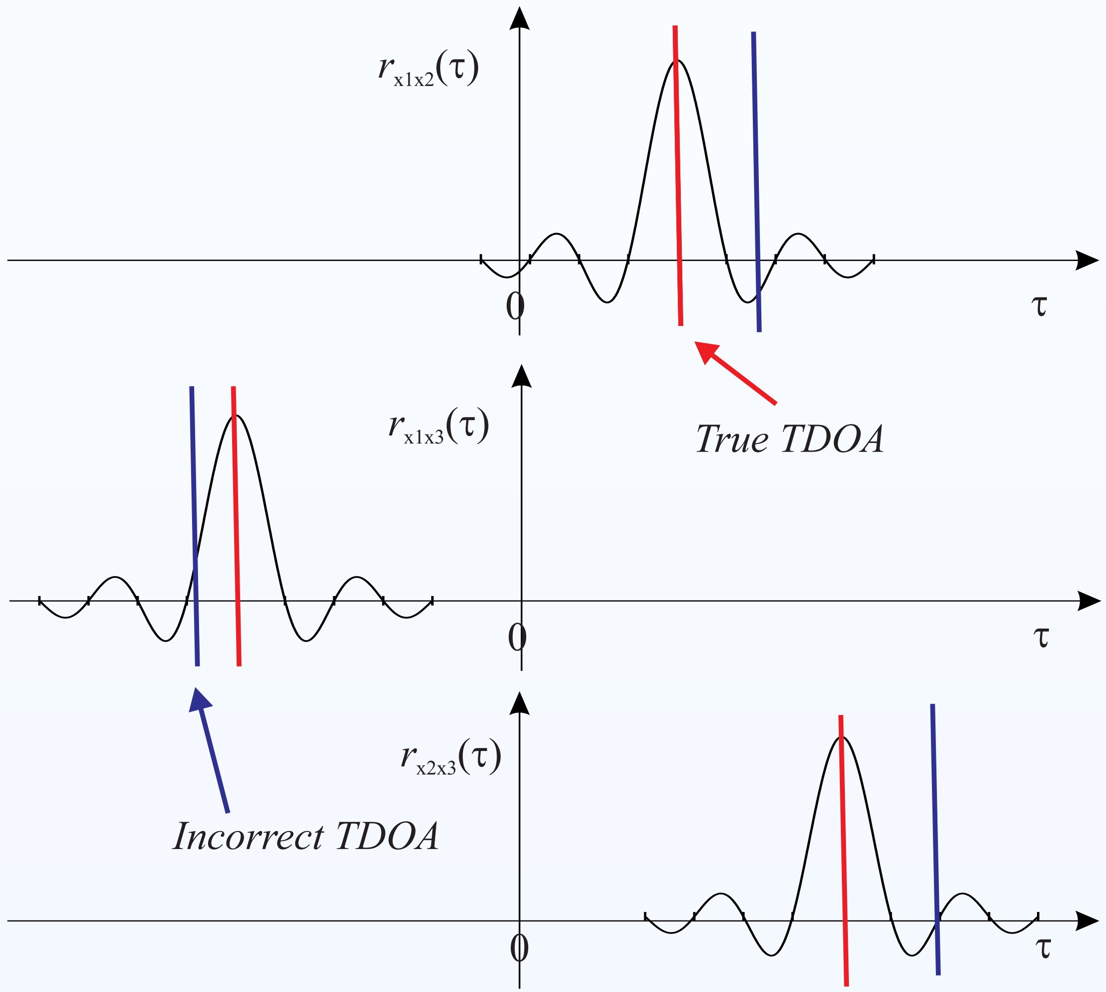
Steered Response Power Function



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

- Show video!

Conceptual Intepretation



GCC-PHAT for different microphone pairs.

$$T(\mathbf{m}_i, \mathbf{m}_j, \hat{\mathbf{x}}_k) = \frac{|\hat{\mathbf{x}}_k - \mathbf{m}_i| - |\hat{\mathbf{x}}_k - \mathbf{m}_j|}{c}$$

DUET Algorithm

The degenerate unmixing estimation technique (DUET) algorithm is an approach to BSS that ties in neatly to ASL. Under certain assumptions and circumstances, it is possible to separate more than two sources using only two microphones.

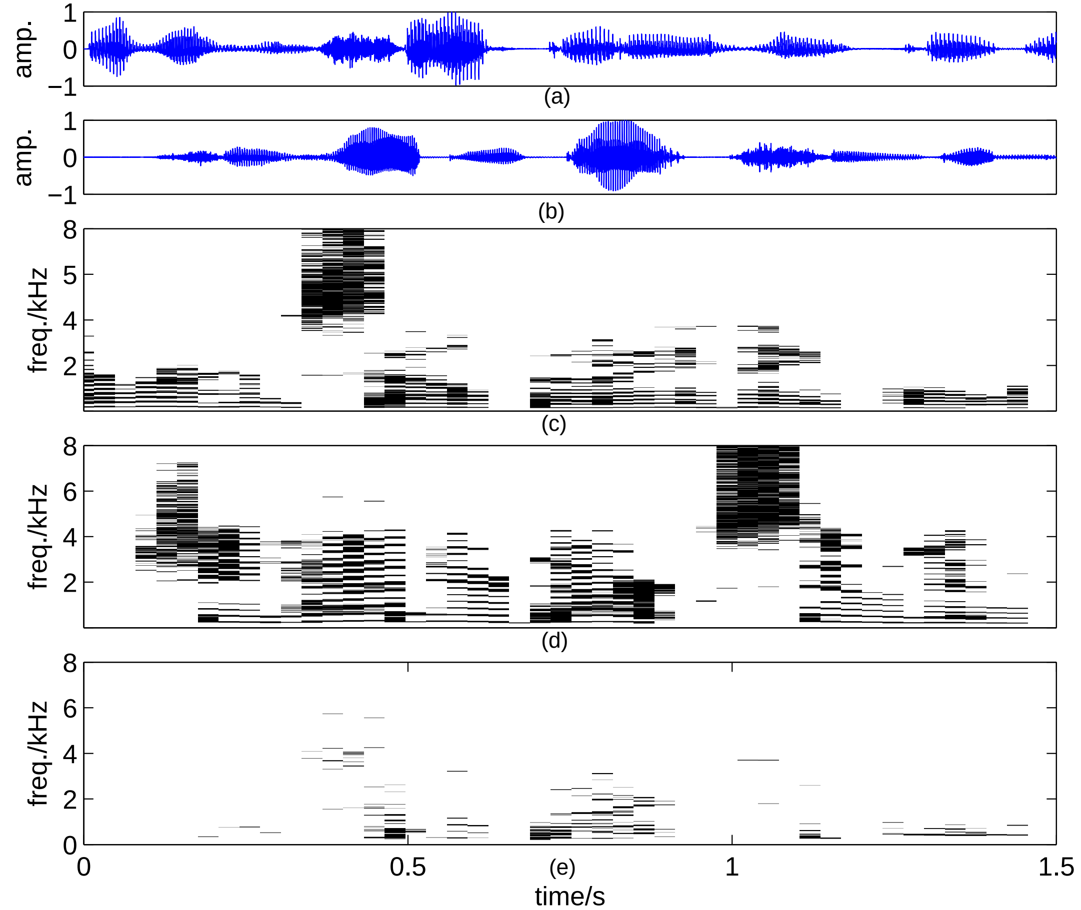
DUET Algorithm

The DUET algorithm is an approach to BSS that ties in neatly to ASL. Under certain assumptions and circumstances, it is possible to separate more than two sources using only two microphones.

- DUET is based on the assumption that for a set of signals $x_k[t]$, their time-frequency representations (TFRs) are predominately non-overlapping. This condition is referred to as W-disjoint orthogonality (WDO):

$$S_p(\omega, t)\, S_q(\omega, t) = 0 \,\forall p \neq q,\, \forall t,\, \omega$$

DUET Algorithm



W-disjoint orthogonality of two speech signals. Original speech signal (a) $s_1[t]$ and (b) $s_2[t]$; corresponding STFTs (c) $|S_1(\omega, t)|$ and (d) $|S_2(\omega, t)|$; (e) product $|S_1(\omega, t)\, S_2(\omega, t)|$.

DUET Algorithm

Consider taking a particular time-frequency (TF)-bin, (ω, t), where source p is known to be active. The two received signals in *that TF-bin* can be written as:

$$X_{ip}(\omega, t) = \alpha_{ip}\, e^{-j\omega\,\tau_{ip}}\, S_p(\omega, t) + B_i(\omega, t)$$

$$X_{jp}(\omega, t) = \alpha_{jp}\, e^{-j\omega\,\tau_{jp}}\, S_p(\omega, t) + B_j(\omega, t)$$

DUET Algorithm

Consider taking a particular TF-bin, $(\omega,\, t)$, where source p is known to be active. The two received signals in *that TF-bin* can be written as:

$$X_{ip}\,(\omega,\, t) = \alpha_{ip}\, e^{-j\omega\, \tau_{ip}}\, S_p\,(\omega,\, t) + B_i\,(\omega,\, t)$$

$$X_{jp}\,(\omega,\, t) = \alpha_{jp}\, e^{-j\omega\, \tau_{jp}}\, S_p\,(\omega,\, t) + B_j\,(\omega,\, t)$$

Taking the ratio and ignoring the noise terms gives:

$$H_{ikp}\,(\omega,\, t) \triangleq \frac{X_{ip}\,(\omega,\, t)}{X_{jp}\,(\omega,\, t)} = \frac{\alpha_{ip}}{\alpha_{jp}}\, e^{-j\omega\tau_{ijp}}$$

DUET Algorithm

Consider taking a particular TF-bin, (ω, t), where source p is known to be active. The two received signals in *that TF-bin* can be written as:

$$X_{ip}(\omega, t) = \alpha_{ip} \, e^{-j\omega \, \tau_{ip}} \, S_p(\omega, t) + B_i(\omega, t)$$
$$X_{jp}(\omega, t) = \alpha_{jp} \, e^{-j\omega \, \tau_{jp}} \, S_p(\omega, t) + B_j(\omega, t)$$

Taking the ratio and ignoring the noise terms gives:

$$H_{ikp}(\omega, t) \triangleq \frac{X_{ip}(\omega, t)}{X_{jp}(\omega, t)} = \frac{\alpha_{ip}}{\alpha_{jp}} \, e^{-j\omega\tau_{ijp}}$$

Hence,

$$\tau_{ijp} = -\frac{1}{\omega} \arg H_{ikp}(\omega, t), \qquad \text{and} \qquad \frac{\alpha_{ip}}{\alpha_{jp}} = |H_{ikp}(\omega, t)|$$

DUET Algorithm

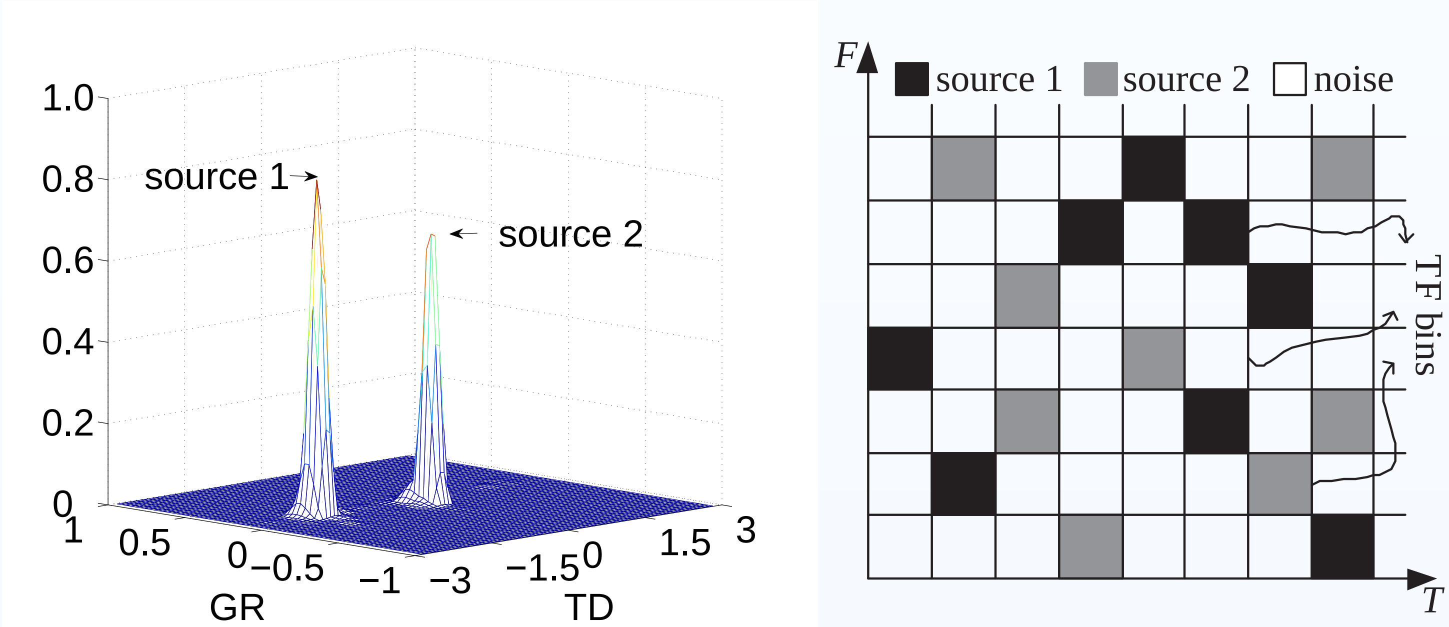


Illustration of the underlying idea in DUET.

DUET Algorithm

This leads to the essentials of the DUET method which are:

1. Construct the TF representation of both mixtures.

2. Take the ratio of the two mixtures and extract local mixing parameter estimates.

DUET Algorithm

This leads to the essentials of the DUET method which are:

1. Construct the TF representation of both mixtures.
2. Take the ratio of the two mixtures and extract local mixing parameter estimates.
3. Combine the set of local mixing parameter estimates into N pairings corresponding to the true mixing parameter pairings.
4. Generate one binary mask for each determined mixing parameter pair corresponding to the TF-bins which yield that particular mixing parameter pair.

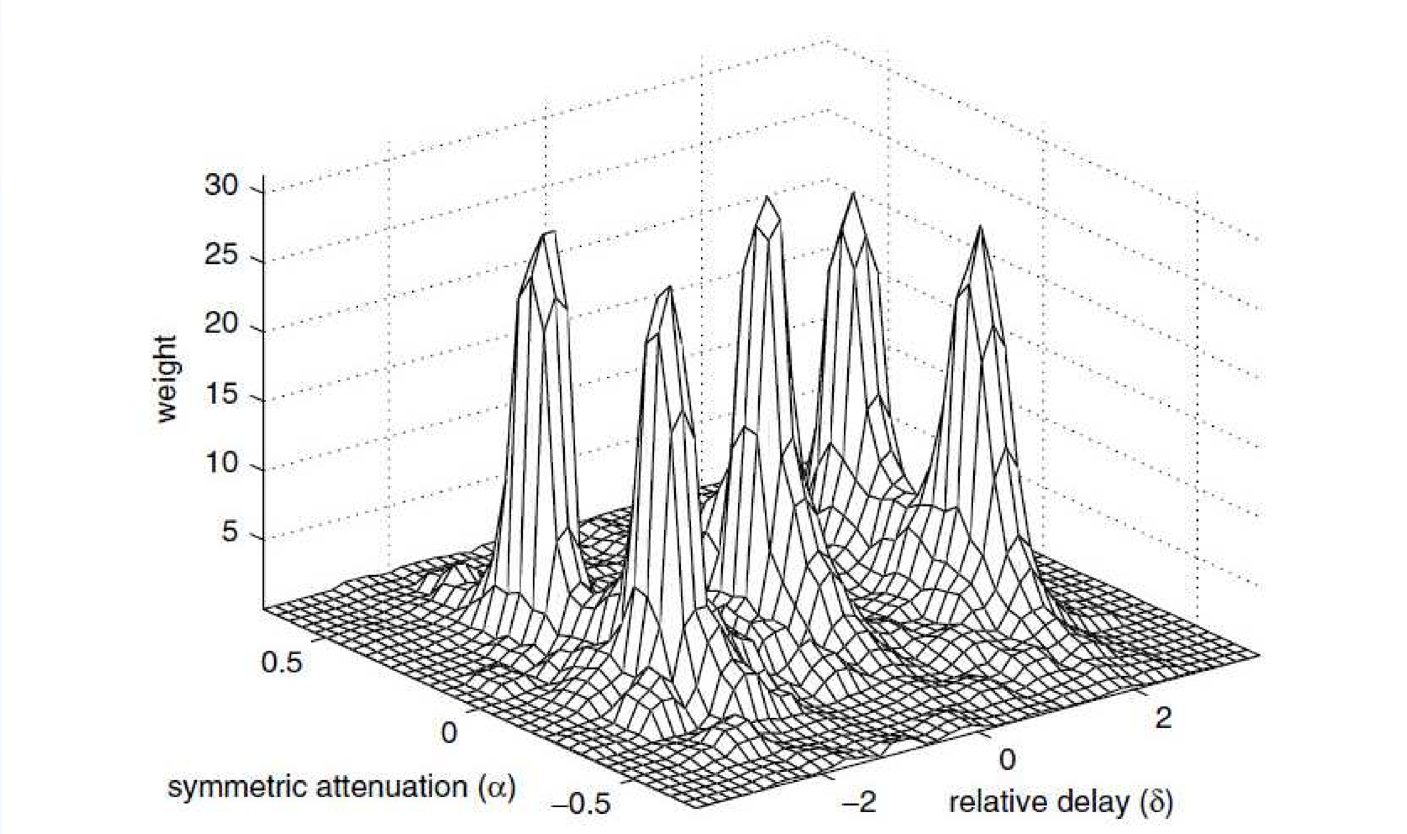
DUET Algorithm

This leads to the essentials of the DUET method which are:

1. Construct the TF representation of both mixtures.
2. Take the ratio of the two mixtures and extract local mixing parameter estimates.
3. Combine the set of local mixing parameter estimates into N pairings corresponding to the true mixing parameter pairings.
4. Generate one binary mask for each determined mixing parameter pair corresponding to the TF-bins which yield that particular mixing parameter pair.
5. Demix the sources by multiplying each mask with one of the mixtures.
6. Return each demixed TFR to the time domain.

DUET Algorithm

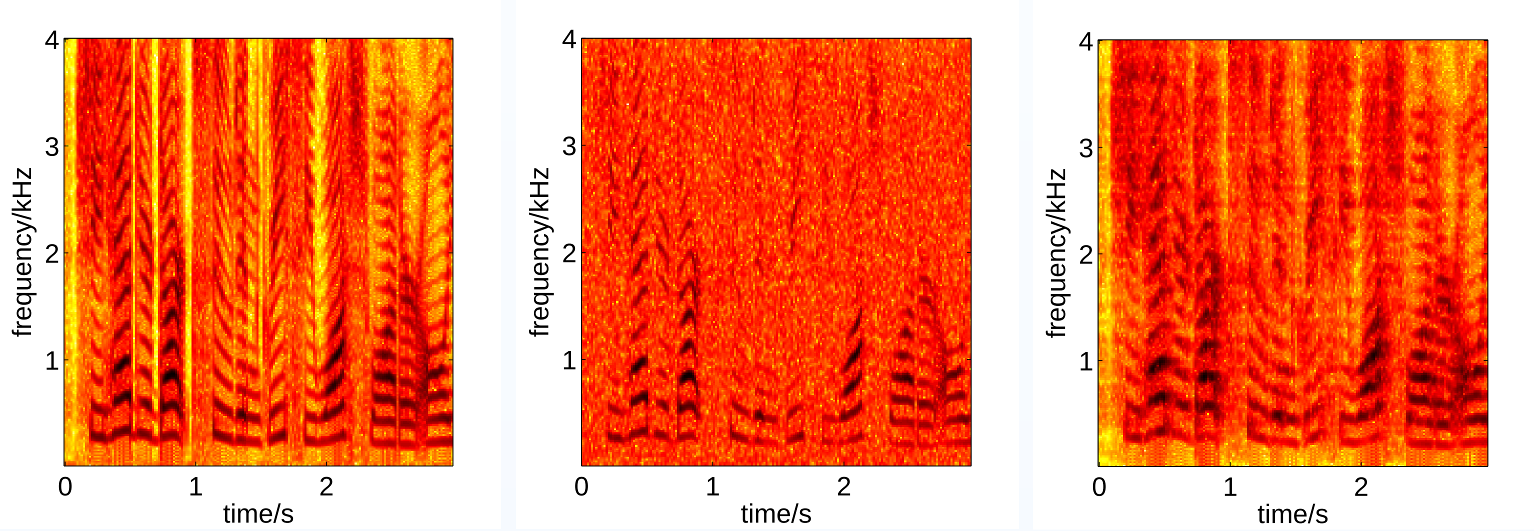
This leads to the essentials of the DUET method which are:



DUET for multiple sources.

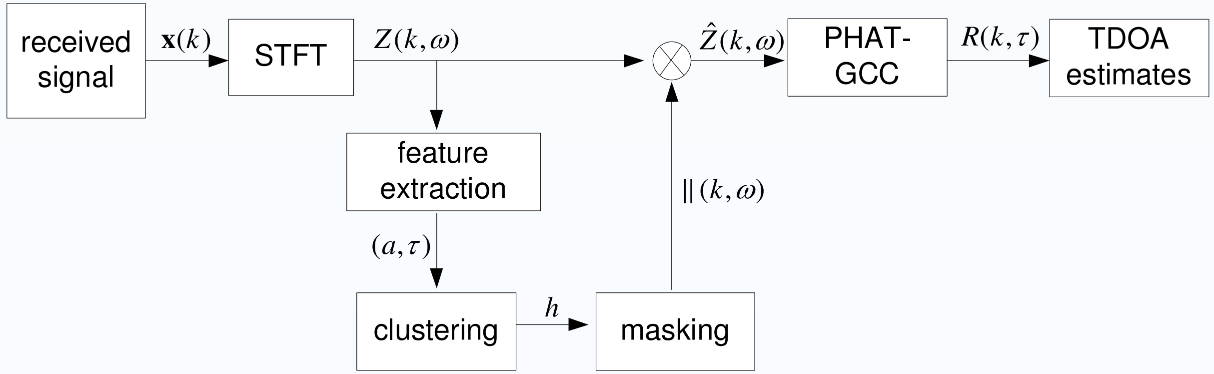
DUET Algorithm

Effect of Reverberation and Noise

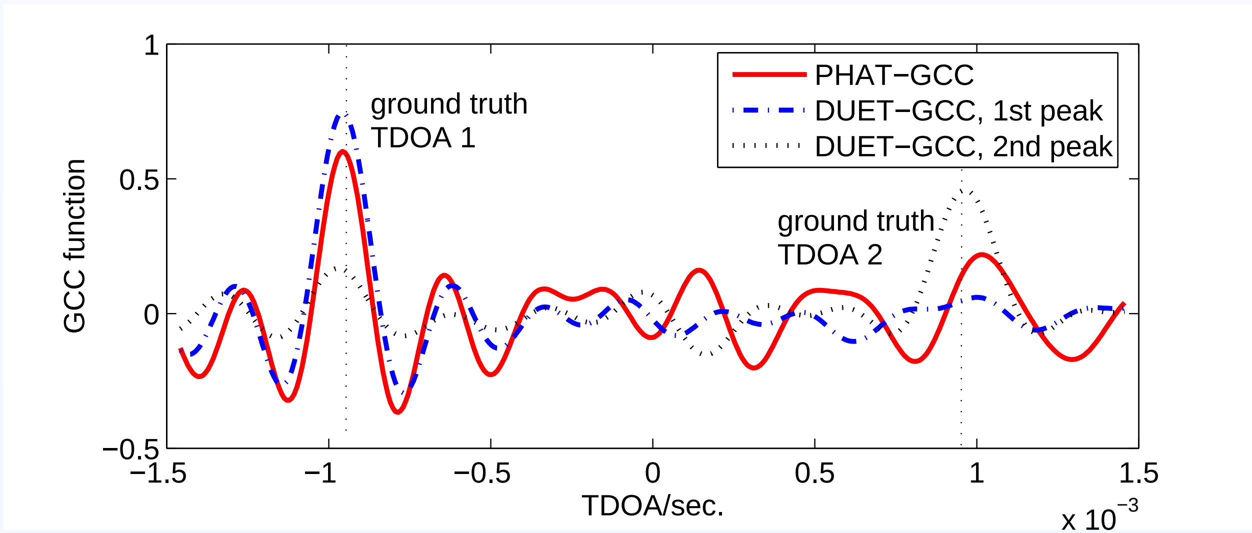


The TFR is very clear in the anechoic environment but smeared around by the reverberation and noise.

Estimating multiple targets



Flow diagram of the DUET-GCC approach. Basically, the speech mixtures are separated by using the DUET in the TF domain, and the PHAT-GCC is then employed for the spectrogram of each source to estimate the TDOAs.



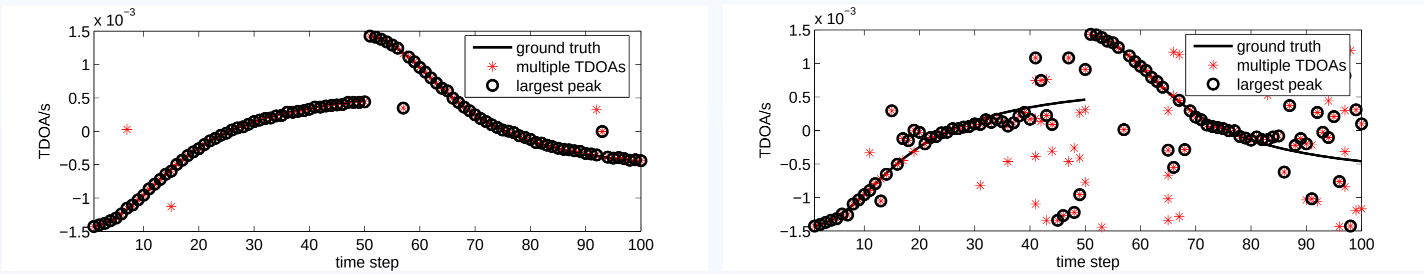
GCC function from DUET approach and traditional PHAT weighting. Two sources are located at $(1.4, 1.2)$m and

Further Topics

- Reduction in complexity of calculating SRP. This includes stochastic region contraction (SRC) and hierarchical searches.

- Multiple-target tracking (see Daniel Clark's Notes)

- Simultaneous (self-)localisation and tracking; estimating sensor and target positions from a moving source.



Acoustic source tracking and localisation.

Further Topics

- Joint ASL and BSS.
- Explicit signal and channel modelling! (None of the material so forth cares whether the signal is speech or music!)
- Application areas such as gunshot localisation; other sensor modalities; diarisation.