Probability and Random Variables; and Classical Estimation Theory

UDRC Summer School, 27th June 2016

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

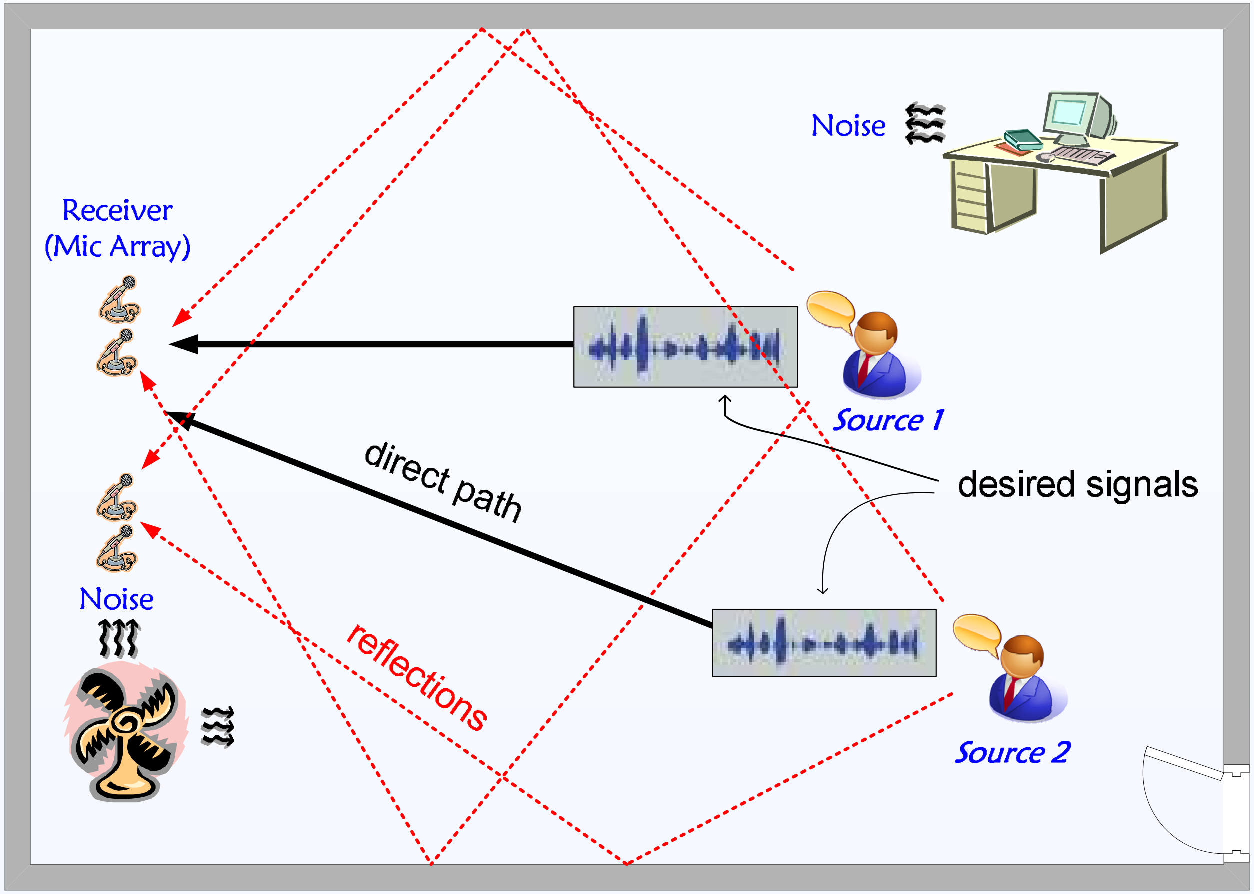
University of Edinburgh

Blank Page

This slide is intentionally left blank.

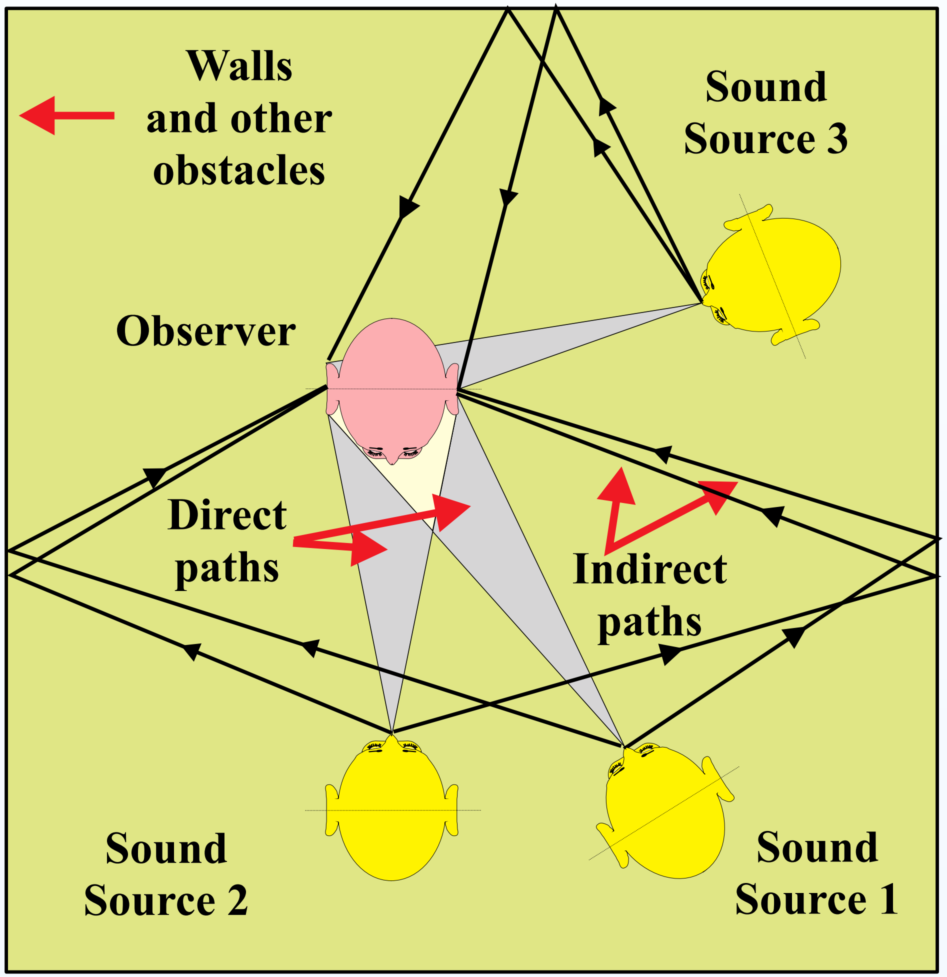
Handout 1
Aims and Objectives

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.

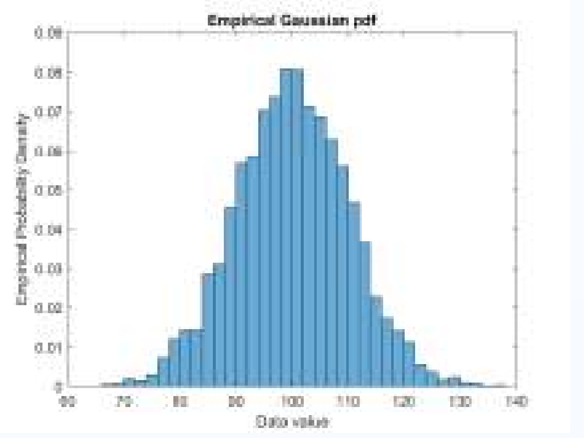
Obtaining the Latest Handouts

- 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 documents published on the USB stick may differ 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: at:

  mailto:james.hopgood@ed.ac.uk

  (Update: The notes are no longer online due to the desire to maintain copyright control on the document.)
- Extended thanks are given to the many MSc students over the past 12 years who have helped proof-read and improve these documents.

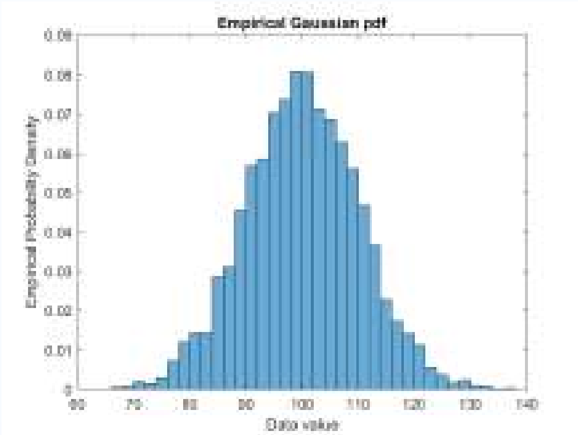
Module Abstract



This topic is covered in two related lecture modules:

1. *Probability, Random Variables, and Estimation Theory*, and
2. *Statistical Signal Processing*,

Module Abstract



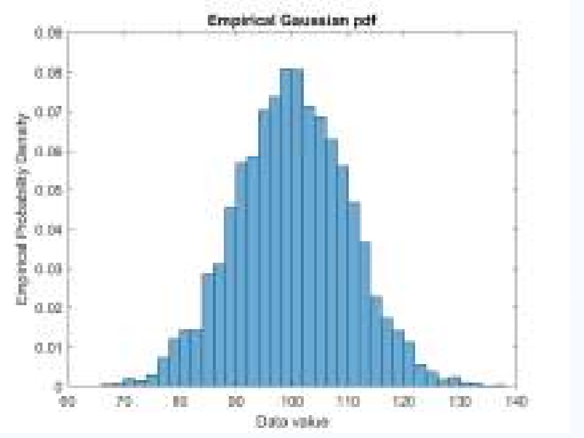
This topic is covered in two related lecture modules:

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

2. *Statistical Signal Processing*,

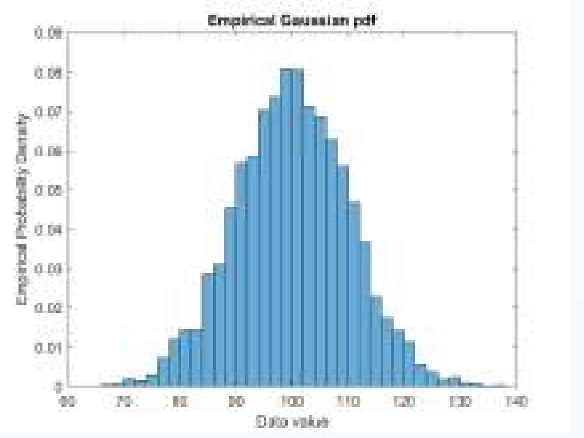
- **Random signals** are extensively used in algorithms, and are:
  - constructively used to model real-world processes;
  - described using *probability and statistics*.

Module Abstract



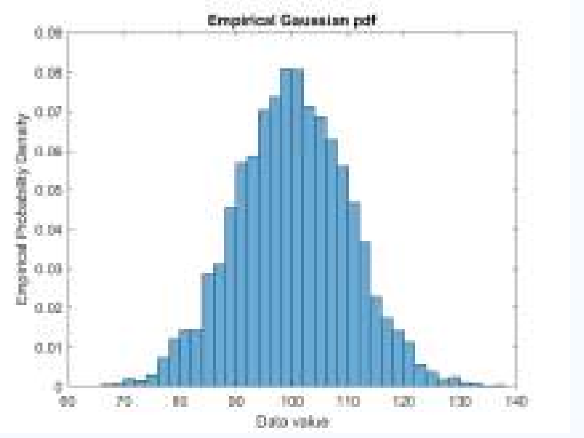
- Their properties are estimated by assumming:
  - an infinite number of observations or data points;
  - time-invariant statistics.

Module Abstract



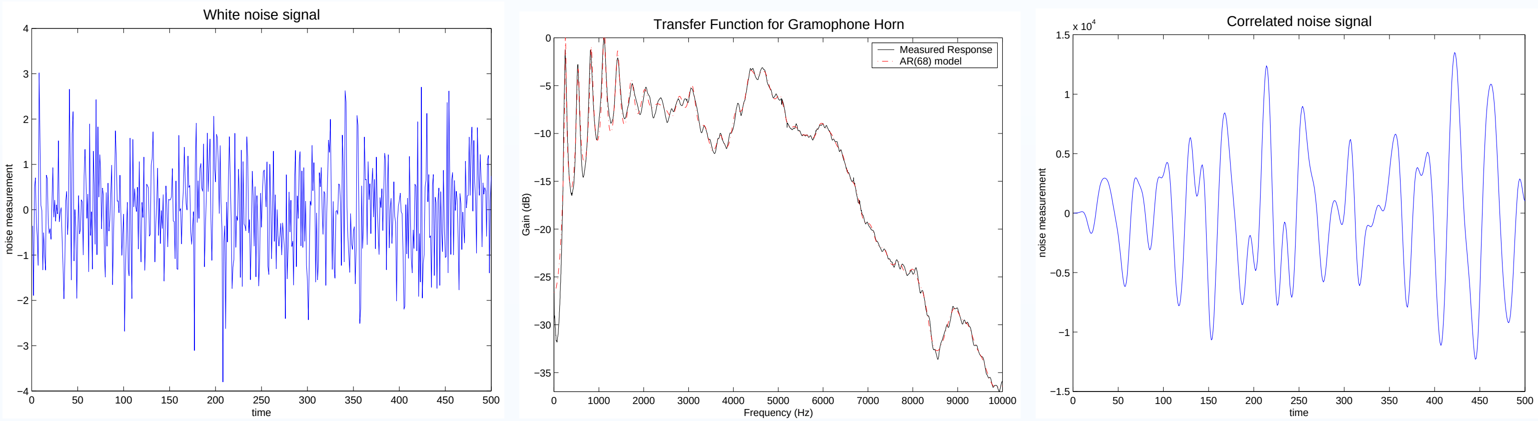
- Their properties are estimated by assumming:
  - an infinite number of observations or data points;
  - time-invariant statistics.
- In practice, these statistics must be estimated from finite-length data signals in noise.

Module Abstract



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

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.

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.

Structure of the Module

These topics are:

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

Structure of the Module

These topics are:

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

Structure of the Module

These topics are:

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

Structure of the Module

These topics 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

These topics 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

These topics 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

These topics 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

These topics are:

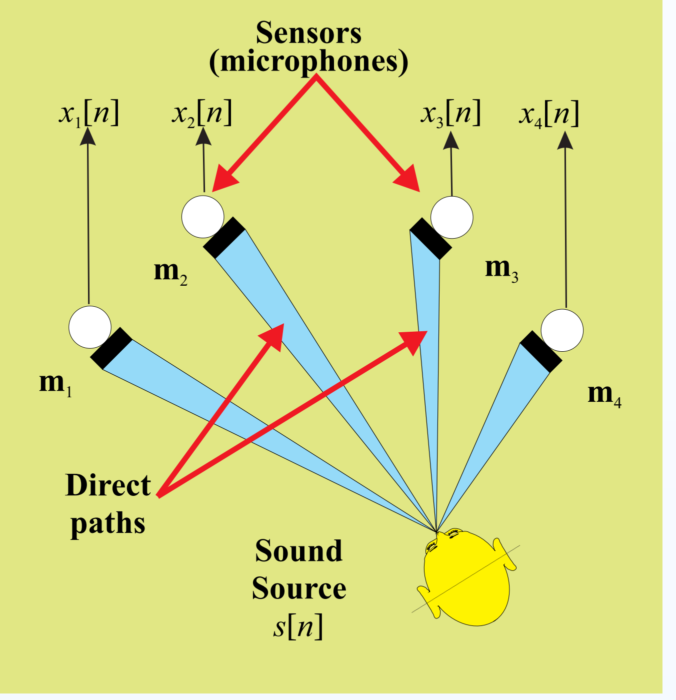
1. review of the fundamentals of **probability theory**;
2. **random variables** and **stochastic processes**;
3. principles of **estimation theory**;
4. **Bayesian estimation** theory;
5. review of **Fourier transforms** and **discrete-time systems**;
6. **linear systems** with stationary random inputs,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

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. 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.

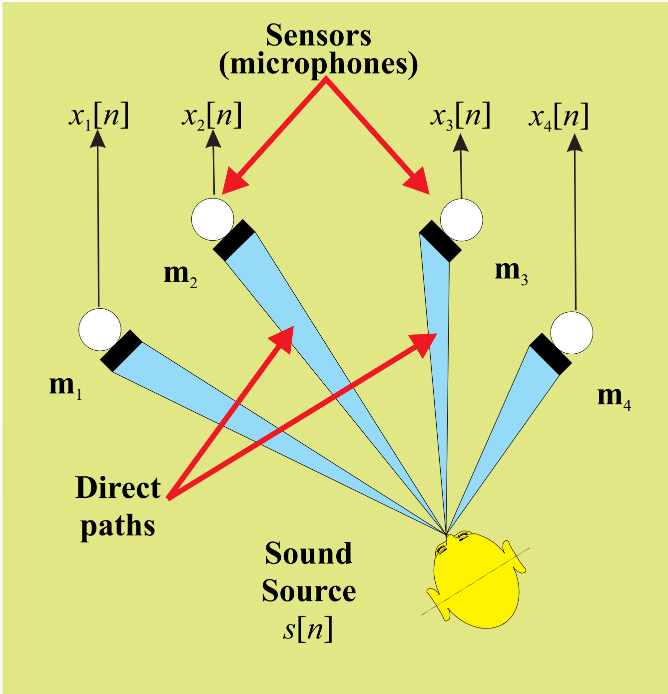
Passive Target Localisation Methodology



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.

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.
- Most 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 sum version of the signal data;

Source Localization Strategies

Existing source localisation methods can loosely be divided into:

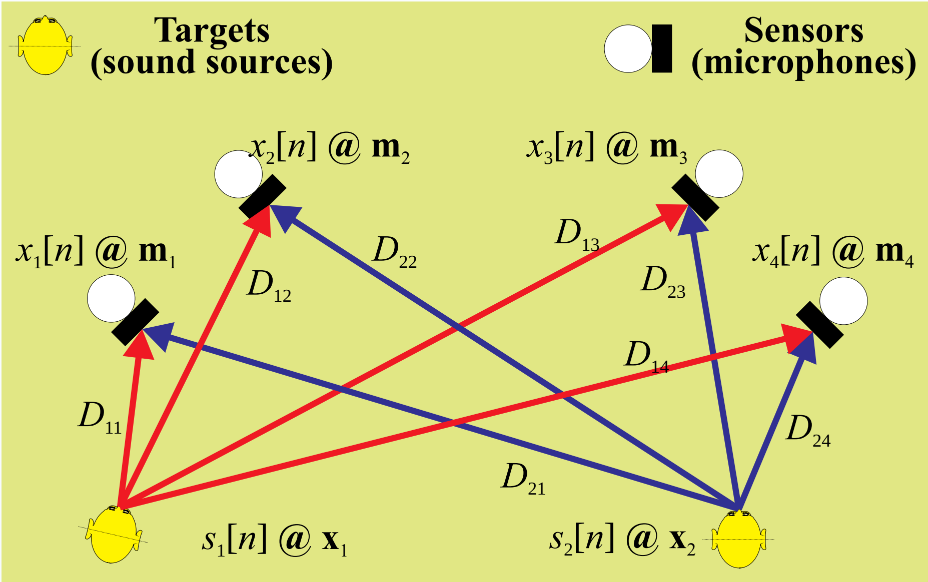
1. those based on maximising the steered response power (SRP) of a beamformer:
   - location estimate derived directly from a filtered, weighted, and sum version of the signal data;
2. techniques adopting high-resolution spectral estimation concepts:
   - any localisation scheme relying upon an application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

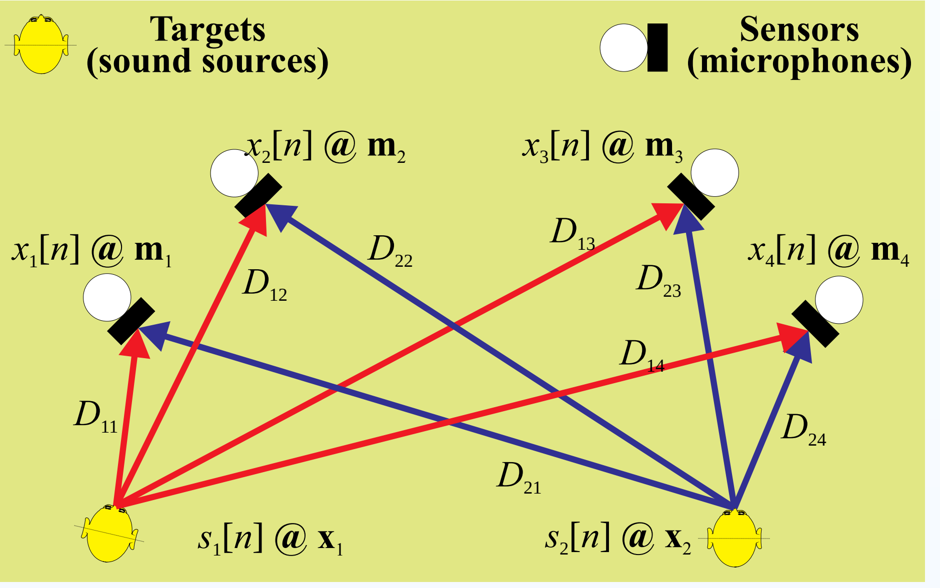


Geometry assuming a free-field model.

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.

- 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.

- 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

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

- The total error as a function of target position

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

where

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

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

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:

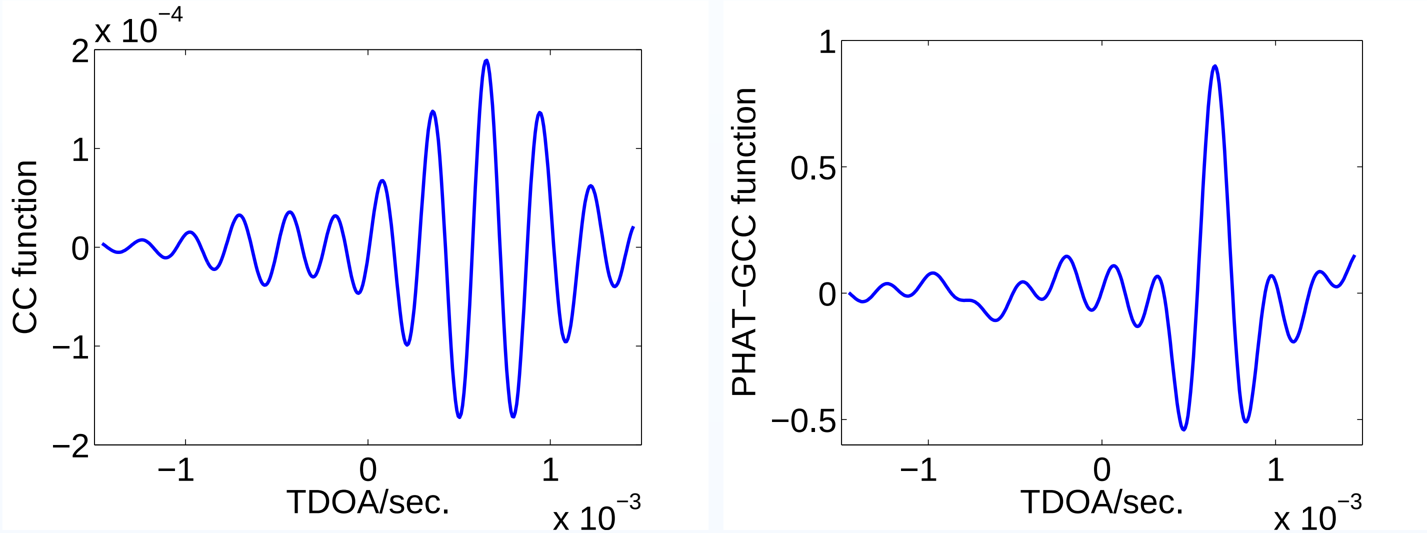
$$\angle P_{x_i x_j}(\omega) = -j\omega\, T\left(\mathbf{m}_i,\, \mathbf{m}_j,\, \mathbf{x}_k\right)$$

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.

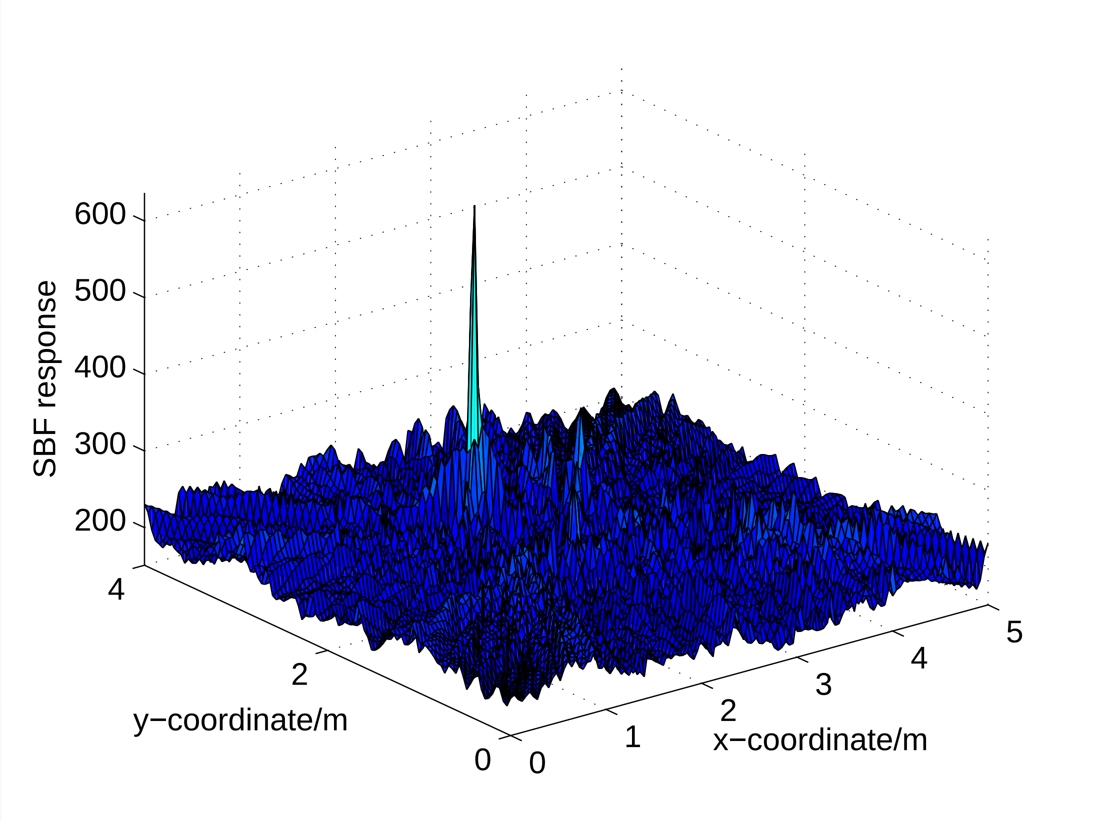
The frequency domain **delay-and-sum beamformer** steered to a spatial position $\hat{\mathbf{x}}_k$ such that $\hat{\tau}_{pk} = |\hat{\mathbf{x}} - \mathbf{m}_p|$:

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

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

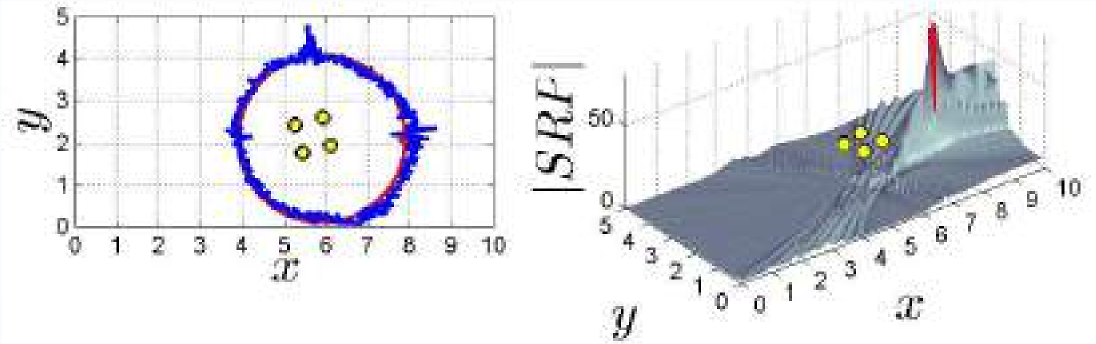
$$\equiv \sum_{p=1}^{N} \sum_{q=1}^{N} r_{x_i x_j}\left[\frac{|\mathbf{x}_k - \mathbf{m}_i| - |\mathbf{x}_k - \mathbf{m}_j|}{c}\right]$$

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!

Handout 2
Probability Theory

Introduction

How many water taxis are there in Venice?

Introduction


How many water taxis are there in Venice?


How does your answer change when you see more taxis?

Introduction

- The theory of probability deals with averages of mass phenomena occurring sequentially or simultaneously;
  - this might include radar detection, signal detection, anomaly detection, parameter estimation, ...
- By considering fundamentals such as the probability of individual events, we can develop a probabilistic framework for analysing signals.
- It is *observed* that certain averages approach a constant value as the number of observations increases; and that this value remains the same if the averages are evaluated over any sub-sequence specified before the experiment is performed.

Introduction

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*.

Note that this interpretation and the language used is all very imprecise.

Classical Definition of Probability

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

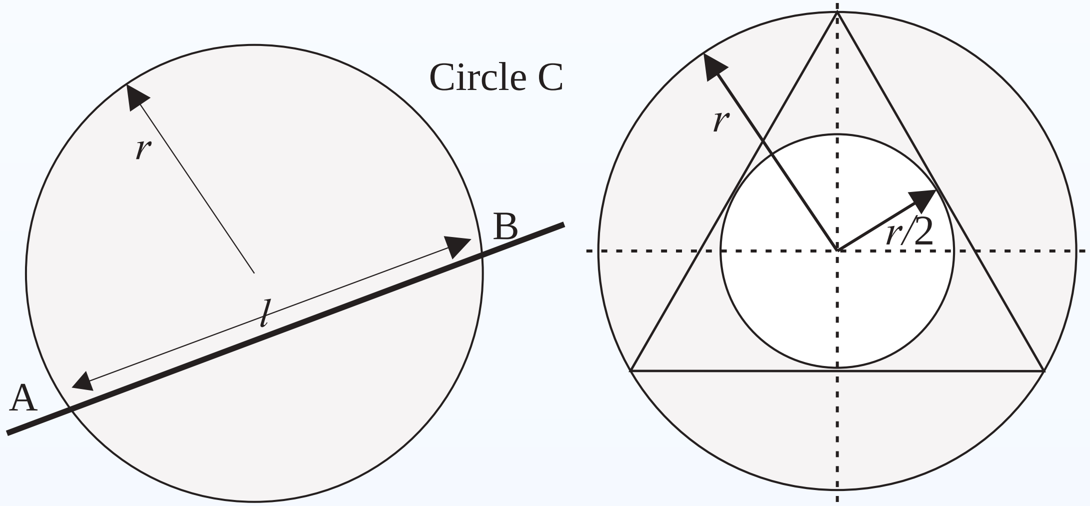
$$\Pr(A) = \frac{N_A}{N}$$

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*.

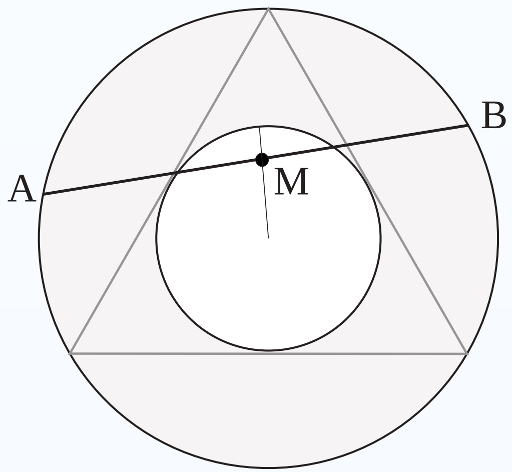
Bertrand's Paradox

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



Bertrand's paradox, problem definition.

Bertrand's Paradox

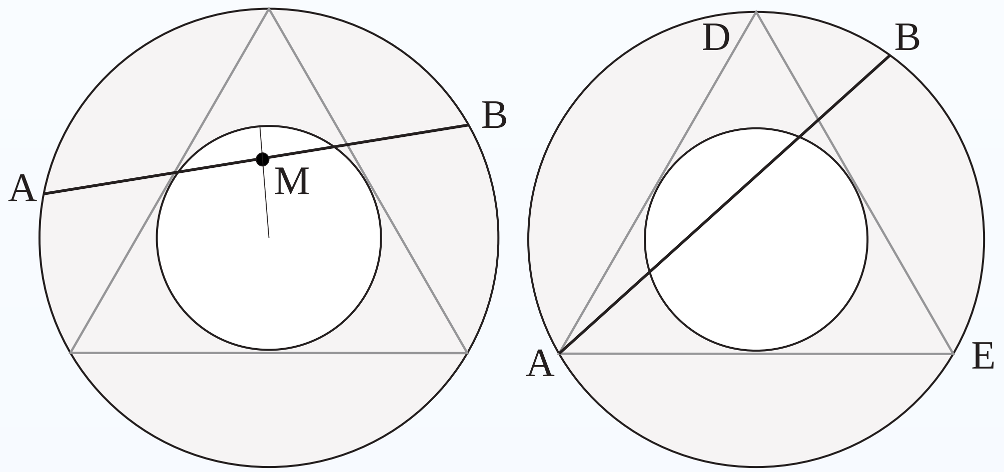


Different selection methods.

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

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

Bertrand's Paradox

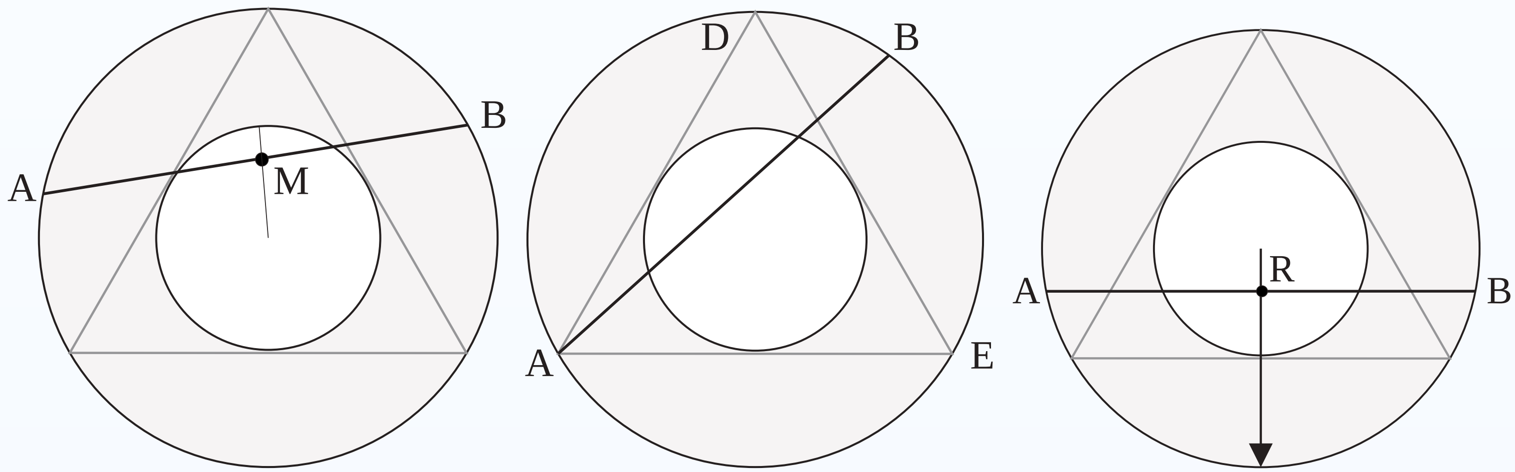


Different selection methods.

1. 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.

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

Bertrand's Paradox



Different selection methods.

1. 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.

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

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

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, as discussed in Bertrand's paradox.

Axiomatic Definition

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 the probability of the certain event equals 1, such that:

$$\Pr(S) = 1$$

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

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

Set Theory

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

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

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

Set Theory

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

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

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

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

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

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

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

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:

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

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

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

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

Set Theory

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

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

As an application of this, note that:

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

Properties of Axiomatic Probability

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

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

Properties of Axiomatic Probability

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

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

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, and therefore:

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

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

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

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

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

Properties of Axiomatic Probability

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

- First, note that

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

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

- Second, note that:

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

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

Properties of Axiomatic Probability

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

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

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

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

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

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 $1/12 \leq \Pr(A\,B) \leq 1/3$.

SOLUTION. Using the sum rule, that:

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

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

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.

Conditional Probability

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

Conditional Probability

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

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

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

provided that n is sufficiently large.

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

Conditional Probability

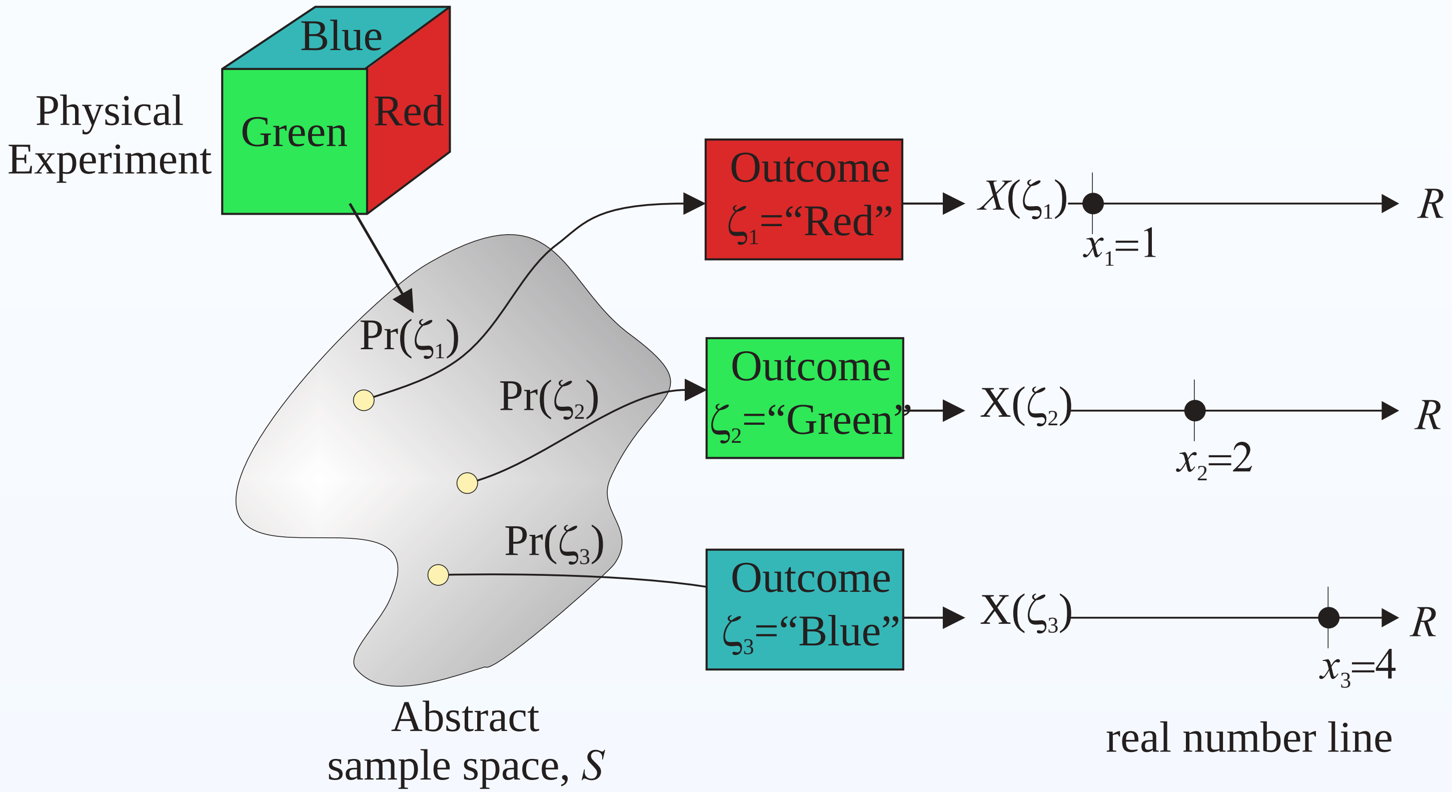
Example (Two Children). A family has two children. What is the probability that both are boys, given that at least one is a boy?

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

Conditional Probability

Handout 3
Scalar Random Variables

Definition



A graphical representation of a random variable.

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 $\{X(\zeta) \le x\}$ is an event in the abstract probability space for every $x \in \mathbb{R}$;

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

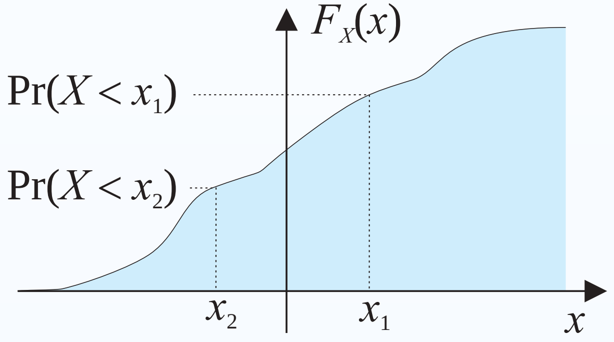
Definition

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:

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

$\bowtie$

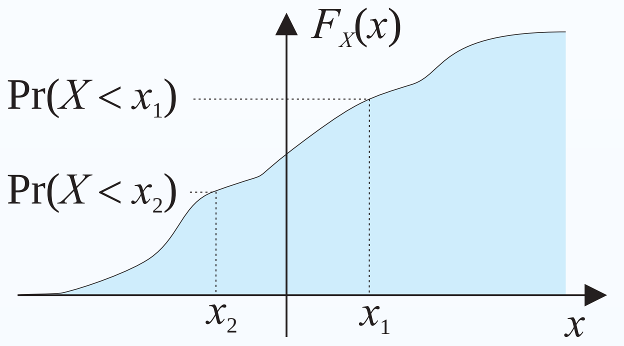
Distribution functions



The cumulative distribution function.

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

Distribution functions

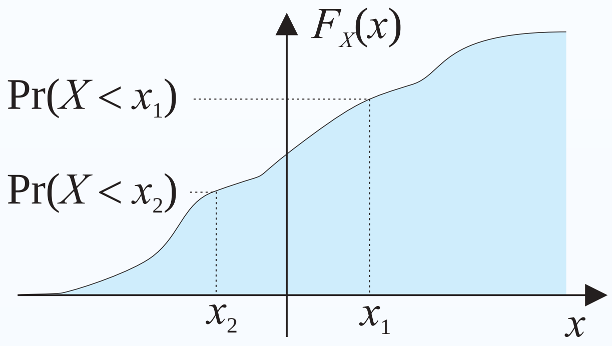


The cumulative distribution function.

- The **probability set function** $\Pr\left(X\left(\zeta\right) \leq x\right)$ is a function of the set $\{X\left(\zeta\right) \leq x\}$, and therefore of the point $x \in \mathbb{R}$.
- This probability is the **cumulative distribution function (cdf)**, $F_X\left(x\right)$ of a **RV** $X\left(\zeta\right)$, and is defined by:

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

Distribution functions

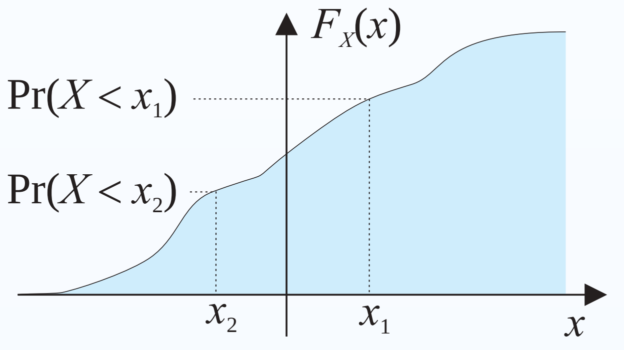


The cumulative distribution function.

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

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

Distribution functions



The cumulative distribution function.

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

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

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

Kolmogorov's Axioms

The events $\{X \le x_1\}$ and $\{x_1 < X \le x_2\}$ are mutually exclusive events. Therefore, their union equals $\{x \le x_2\}$, and therefore:

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

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

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

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

Density functions

- The **probability density function (pdf)**, $f_X(x)$ of a **RV** $X(\zeta)$, is defined as a formal derivative:

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

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

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

Density functions

- The **probability density function (pdf)**, $f_X(x)$ of a **RV** $X(\zeta)$, is defined as a formal derivative:

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

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

- The **probability density function (pdf)**, $f_X(x)$ of a **RV** $X(\zeta)$, is defined as a formal derivative:

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

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

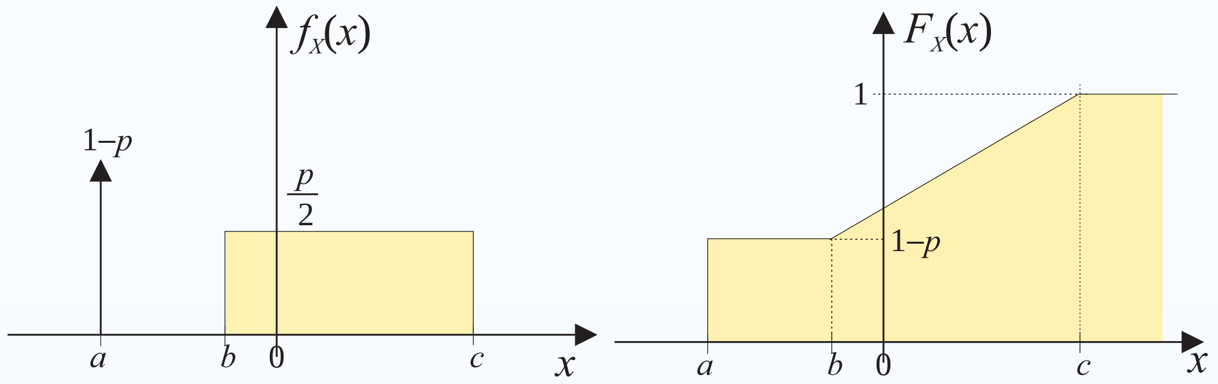
$$f_X(x)\,\Delta x \approx F_X(x+\Delta x) - F_X(x) \approx \Pr\left(x < X(\zeta) \leq x + \Delta x\right)$$

- It directly follows that:

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

- For discrete-valued **RV**, use the **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



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

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:

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

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

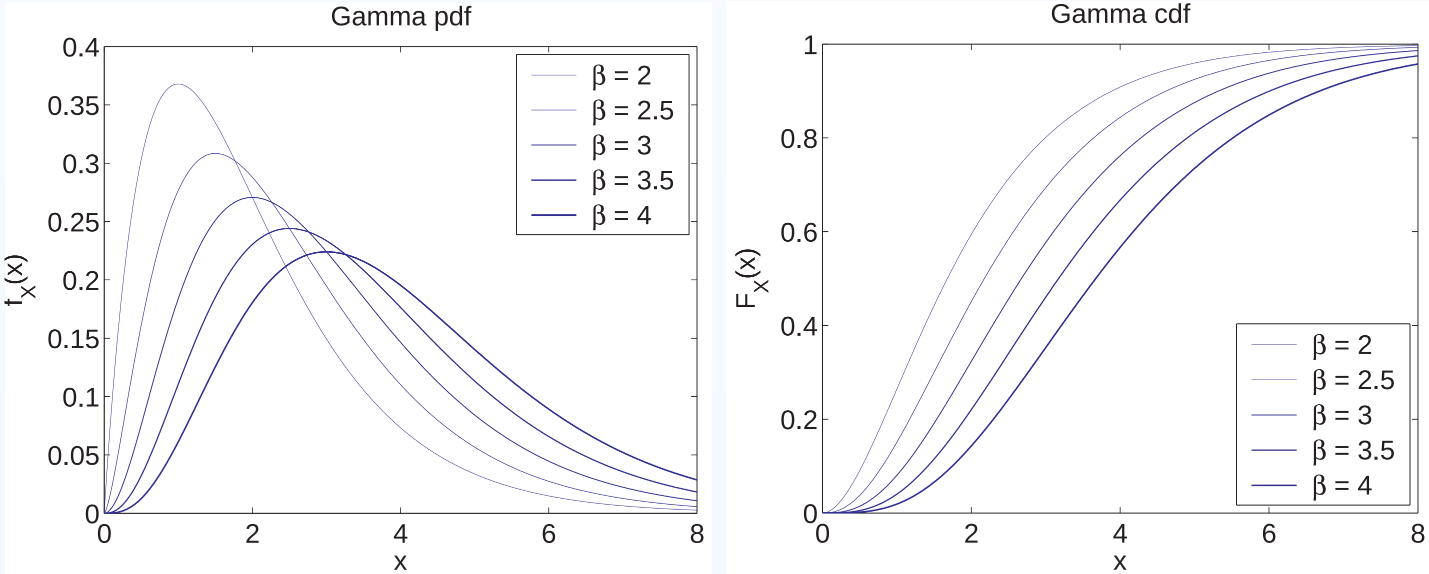
$$f_X(x) = \frac{\beta}{\pi} \frac{1}{(x-\mu_X)^2 + \beta^2}$$

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

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

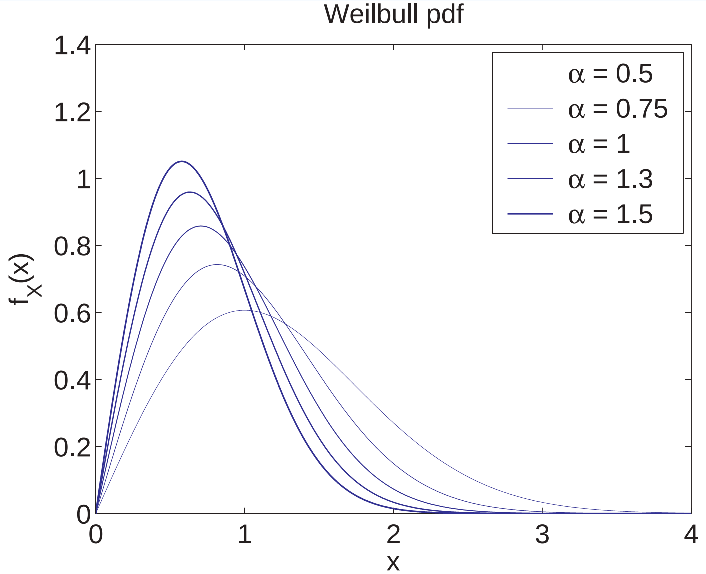


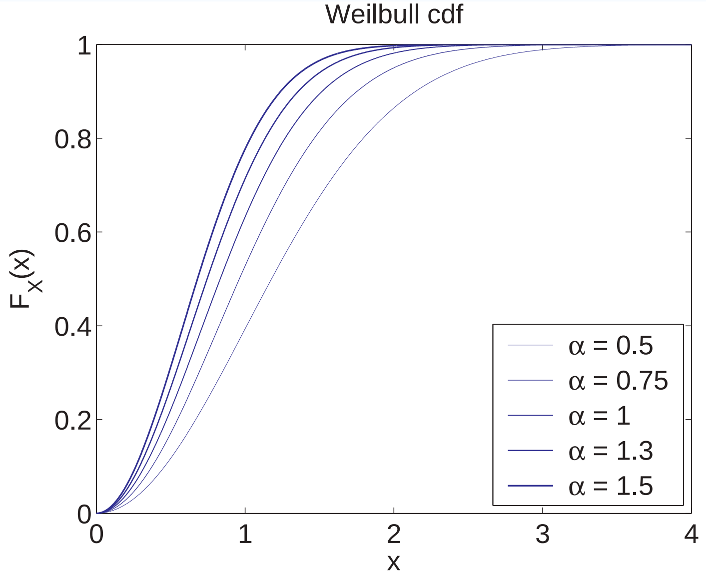
The Gamma density and distribution functions, for the case when $\alpha = 1$ and for various values of β.

Common Continuous RVs

Weibull distribution

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

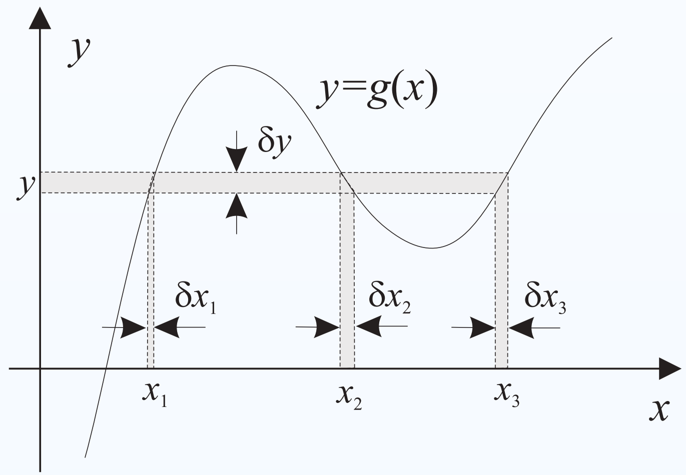




The Weibull density and distribution functions, for the case when $\alpha = 1$, and for various values of the parameter β.

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 $f_Y(y)$?



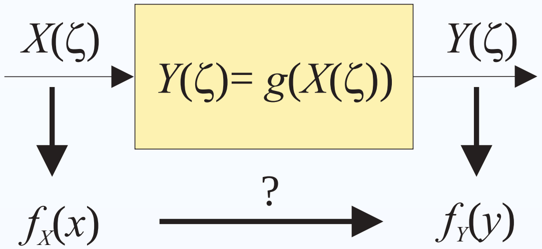
The mapping $y = g(x)$, and the effect of the mapping on intervals.

Aims and Objectives
Probability Theory
Scalar Random Variables
- Definition
- Distribution functions
- Kolmogorov's Axioms
- Density functions
- Properties: Distributions and Densities
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Higher-order statistics

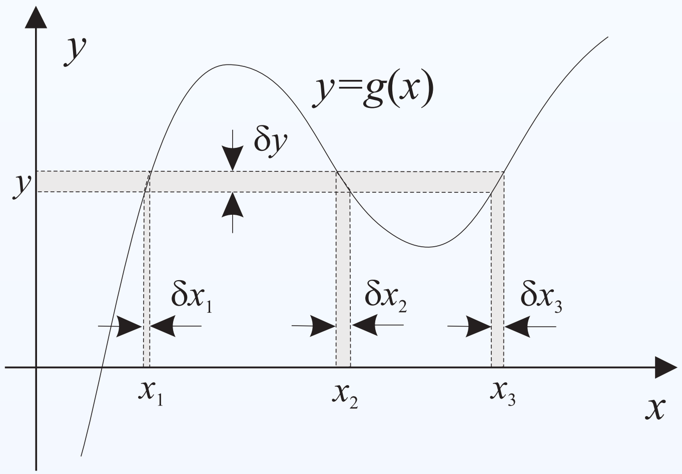
Multiple Random Variables
Estimation Theory
MonteCarlo
Passive Target Localisation

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 $f_Y(y)$?

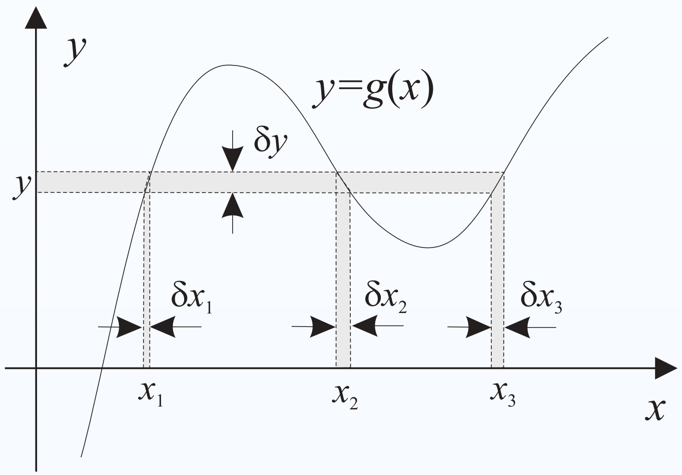


The mapping $y = g(x)$.



The mapping $y = g(x)$, and the effect of the mapping on intervals.

Probability transformation rule



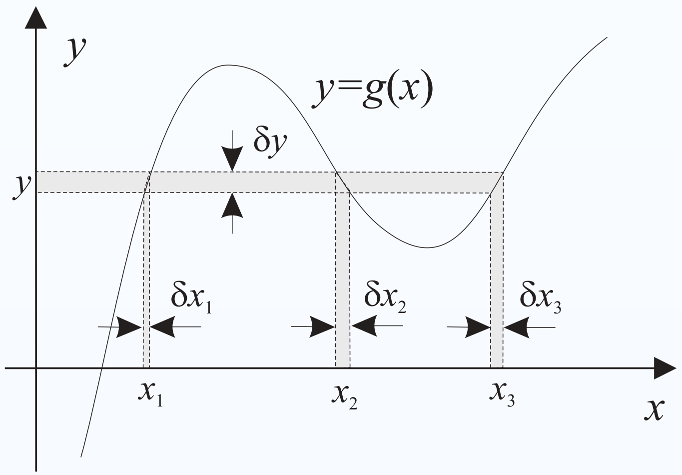
The mapping $y = g(x)$, and the effect of the mapping on intervals.

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

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

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

Probability transformation rule



The mapping $y = g(x)$, and the effect of the mapping on intervals.

Theorem (Probability transformation rule). 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 $Y(\zeta)$ in terms of the pdf of $X(\zeta)$ is given by:

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

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:

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

$\square$

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 $g'(x) = e^x = y$.

Hence, it follows:

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

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

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

Expectations

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

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

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

Expectations

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

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

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:

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

where the order of integration and summation have been interchanged, and the sifting-property applied.

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:

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

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:

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

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

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

Moments

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

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

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

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:

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

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

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 $X\left(\zeta\right)$: the m-th moment of $X\left(\zeta\right)$ 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: $\mathbb{E}\left[X^m(\zeta)\right] \neq \mathbb{E}^m\left[X\left(\zeta\right)\right]$.

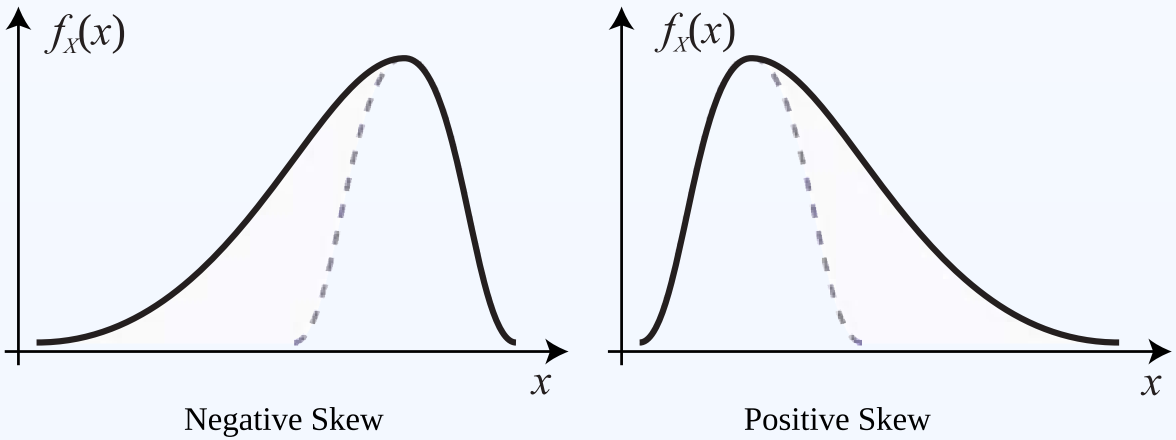
Higher-order statistics

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.



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.

Handout 4
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 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 $\mathbf{X}(\zeta) \le \mathbf{x}$ is equivalent to the event $\{X_n(\zeta) \le x_n,\, n \in \mathcal{N}\}$.

Distribution and Density Functions

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

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

Distribution and Density Functions

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

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

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

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

Distribution and Density Functions

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

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

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

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

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

- Properties of **joint-cdf**:

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

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

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

Distribution and Density Functions

- Properties of **joint-cdf**:

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

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

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

- Properties of **joint-pdfs**:

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

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

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

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:

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

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:

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

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

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

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

$\square$

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:

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

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

$\square$

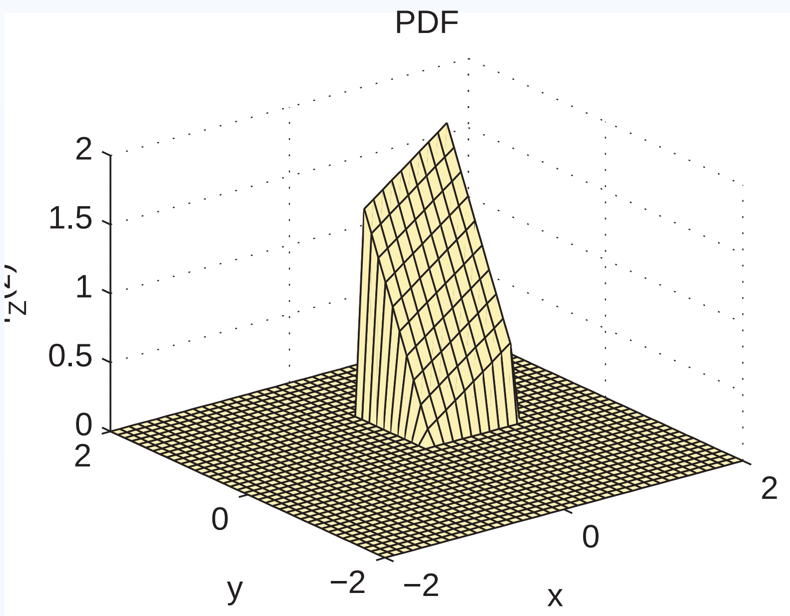
Distribution and Density Functions

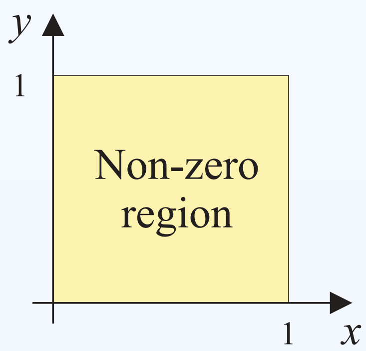
Example ([Therrien:1992, Example 2.1, Page 20]).

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

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

SOLUTION. The pdf is shown here:





Region of support for pdf.

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:

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

$\square$

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:

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

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

$\square$

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:

$$F_{\mathbf{Z}}(\mathbf{z}) = \int_{-\infty}^{\mathbf{z}} f_{\mathbf{Z}}(\bar{\mathbf{z}})\, d\bar{\mathbf{z}} = \int_0^y \int_0^x \frac{1}{2}(\bar{x}+3\bar{y})\, d\bar{x}\, d\bar{y}$$
$$= \int_0^y \frac{1}{2}\left(\frac{x^2}{2} + 3x\bar{y}\right) d\bar{y} = \frac{1}{2}\left(\frac{x^2}{2}y + \frac{3xy^2}{2}\right) = \frac{xy}{4}(x+3y)$$

Finally, if $x > 1$ or $y > 1$, the upper limit of integration for the corresponding variable becomes equal to 1.

Distribution and Density Functions

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

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

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

SOLUTION. Hence, in summary, it follows:

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

$\square$

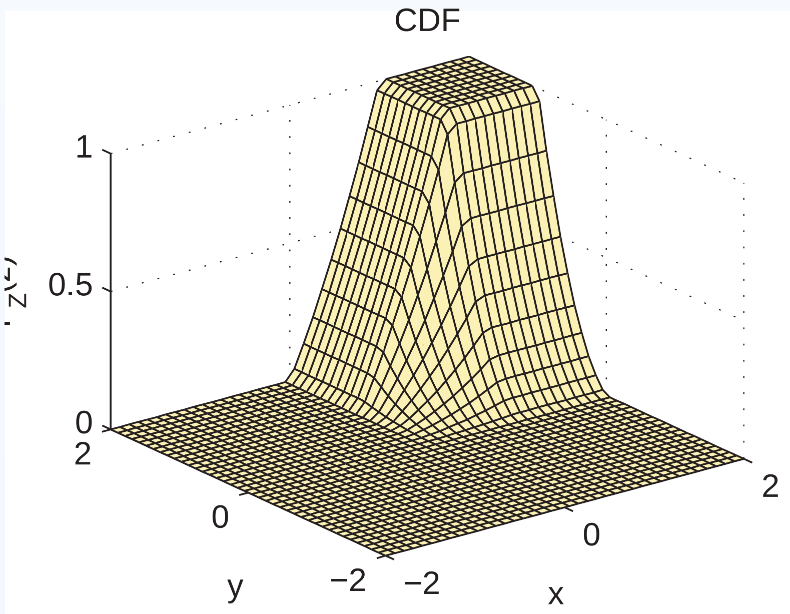
Distribution and Density Functions

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

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

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

SOLUTION. The cdf is plotted here:



A plot of the cumulative distribution function.