



Probability and Random Variables; and Classical Estimation Theory

UDRC Summer School, 20th July 2015

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



Blank Page

This slide is intentionally left blank.

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

Handout 1

Aims and Objectives

Source Signal
e.g. Clean Speech



Observed Signal
e.g. Reverberant Speech



Aims and Objectives

● Module Abstract

- Introduction and Overview
- Description and Learning Outcomes
- Structure of the Module

Probability Theory

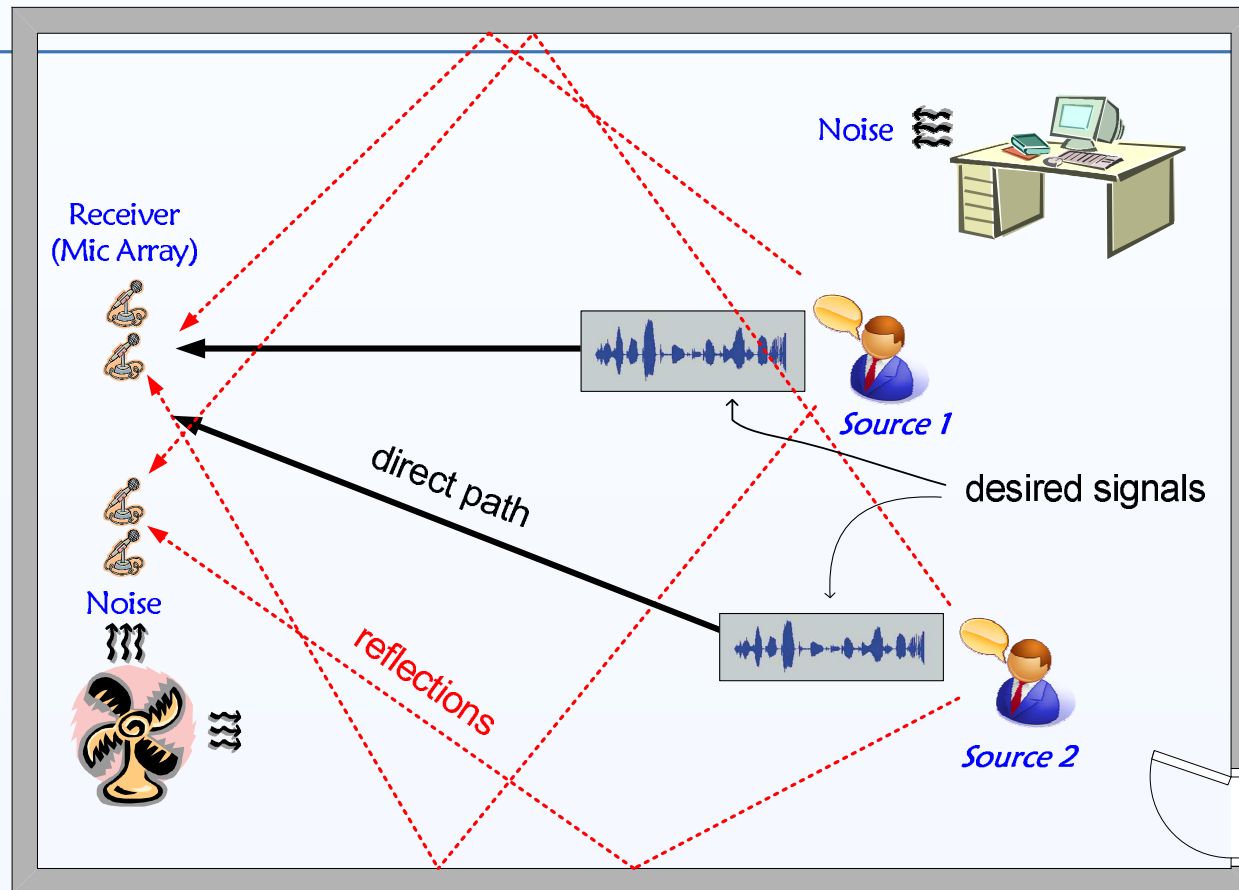
Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

Obtaining the Latest Version of these Hand-outs



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



Aims and Objectives

● Module Abstract

- Introduction and Overview
- Description and Learning Outcomes
- Structure of the Module

Probability Theory

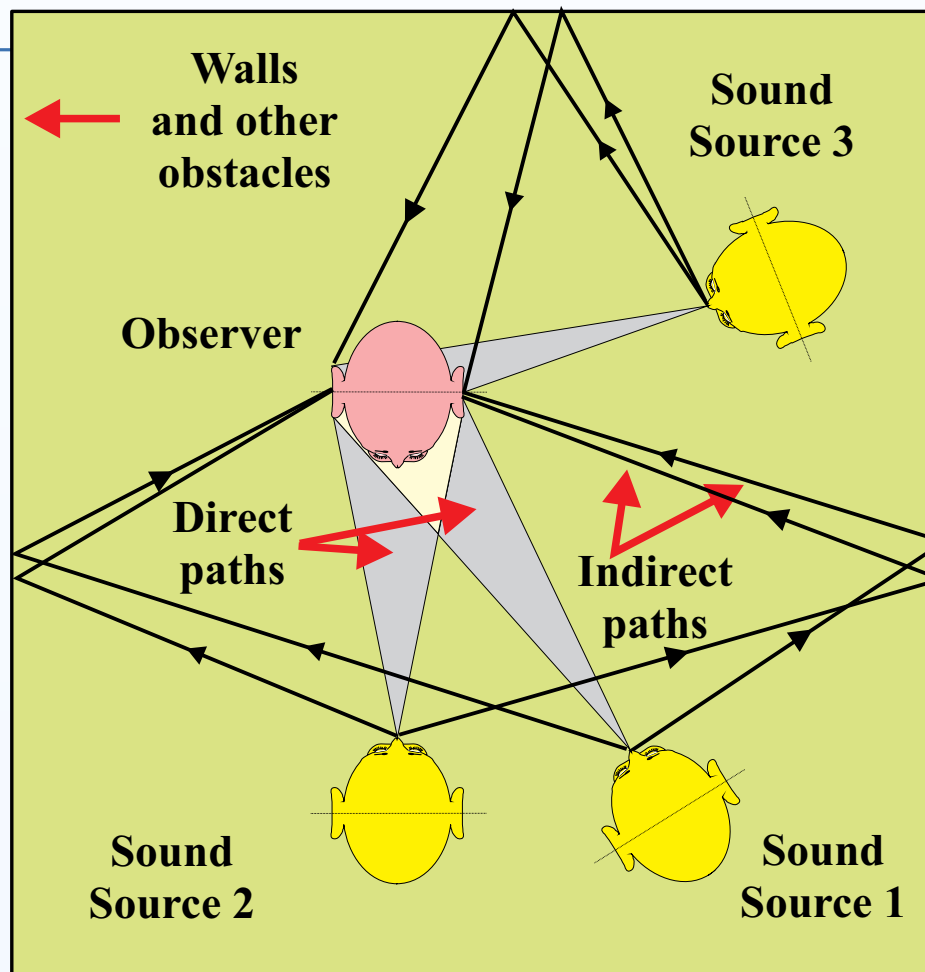
Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

Obtaining the Latest Version of these Hand-outs



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



Aims and Objectives

● Module Abstract

- Introduction and Overview
- Description and Learning Outcomes
- Structure of the Module

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

Obtaining the Latest Version of these Hand-outs

- 📌 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 found online and downloaded at:

<http://www.mod-udrc.org/events/2015-summer-school>

- 📌 Extended thanks are given to the many MSc students over the past 11 years who have helped proof-read and improve these documents.



Aims and Objectives

● Module Abstract

- Introduction and Overview
- Description and Learning Outcomes
- Structure of the Module

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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*.



Aims and Objectives

- Module Abstract
- **Introduction and Overview**
- Description and Learning Outcomes
- Structure of the Module

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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



Aims and Objectives

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

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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*.
- Their properties are estimated by assuming:
 - 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**.



Aims and Objectives

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

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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*.
- Their properties are estimated by assuming:
 - 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**.



Introduction and Overview

Aims and Objectives

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

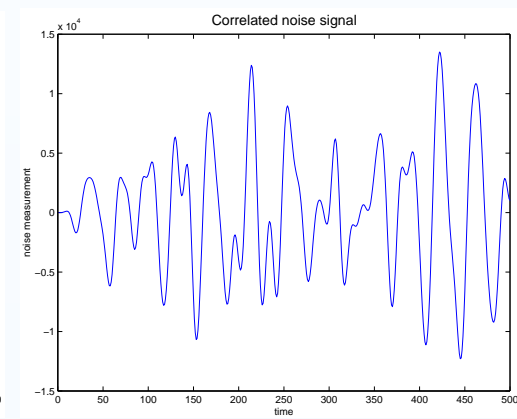
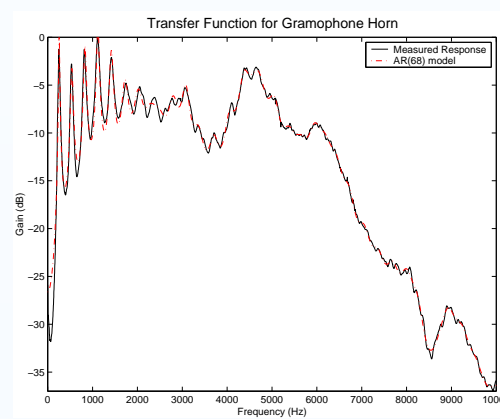
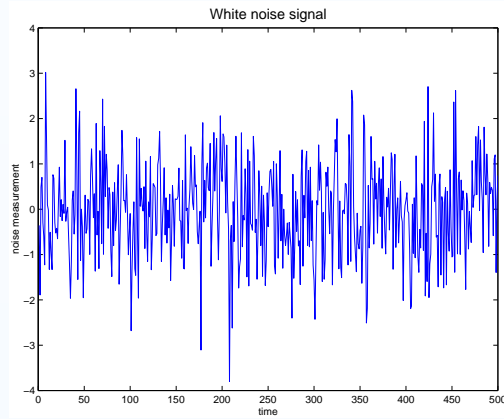
Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo



Source Signal
e.g. Clean Speech

Channel
e.g. Room Acoustics

Observed Signal
e.g. Reverberant Speech

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

Blind deconvolution requires statistical signal processing.

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.



Aims and Objectives

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

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

Description and Learning Outcomes

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



Aims and Objectives

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

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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:

1. acquired sufficient expertise in this area to understand and implement **spectral estimation**, **signal modelling**, **parameter estimation**, and **adaptive filtering** techniques;



Aims and Objectives

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

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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:

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**.



Aims and Objectives

- Module Abstract
- Introduction and Overview
- Description and Learning Outcomes
- **Structure of the Module**

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

Structure of the Module

These topics are:

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



Aims and Objectives

- Module Abstract
- Introduction and Overview
- Description and Learning Outcomes
- **Structure of the Module**

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

Structure of the Module

These topics are:

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



Aims and Objectives

- Module Abstract
- Introduction and Overview
- Description and Learning Outcomes
- **Structure of the Module**

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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**;



Aims and Objectives

- Module Abstract
- Introduction and Overview
- Description and Learning Outcomes
- **Structure of the Module**

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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**;



Aims and Objectives

- Module Abstract
- Introduction and Overview
- Description and Learning Outcomes
- **Structure of the Module**

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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**;



Aims and Objectives

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

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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**;



Aims and Objectives

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

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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**;



Aims and Objectives

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

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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.

Handout 2

Probability Theory



Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- **Bertrand's Paradox**
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- Set Theory
- Properties of Axiomatic Probability
- Countable Spaces
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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, ...
- It is *observed* that certain averages approach a constant value as the number of observations increases, and this value remains the same if the averages are evaluated over any subsequence specified before the experiment is performed.



Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- **Bertrand's Paradox**
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- Set Theory
- Properties of Axiomatic Probability
- Countable Spaces
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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.



Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- **Bertrand's Paradox**
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- Set Theory
- Properties of Axiomatic Probability
- Countable Spaces
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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*.



Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- **Bertrand's Paradox**
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- Set Theory
- Properties of Axiomatic Probability
- Countable Spaces
- The Real Line
- Conditional Probability

Scalar Random Variables

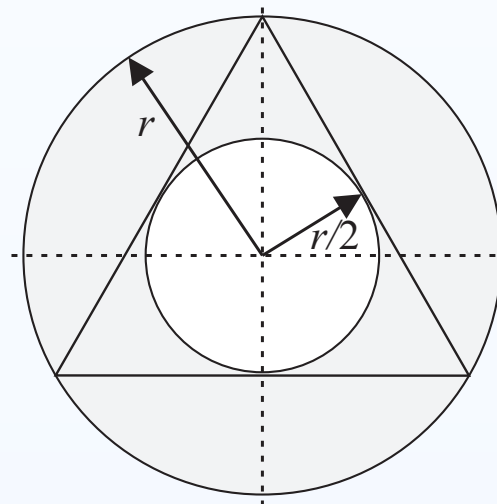
Multiple Random Variables

Estimation Theory

MonteCarlo

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.



Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- Set Theory
- Properties of Axiomatic Probability
- Countable Spaces
- The Real Line
- Conditional Probability

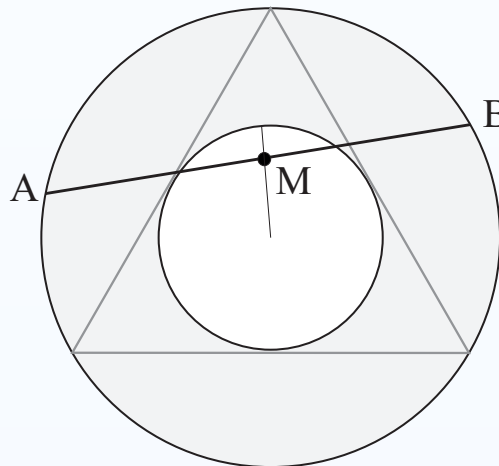
Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

Bertrand's Paradox



Different selection methods.

1. In the **random midpoints** method, a cord is selected by choosing a point M anywhere in the circle, an end-point A on the circumference of the circle, and constructing a chord AB through these chosen points.

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



Bertrand's Paradox

Aims and Objectives

Probability Theory

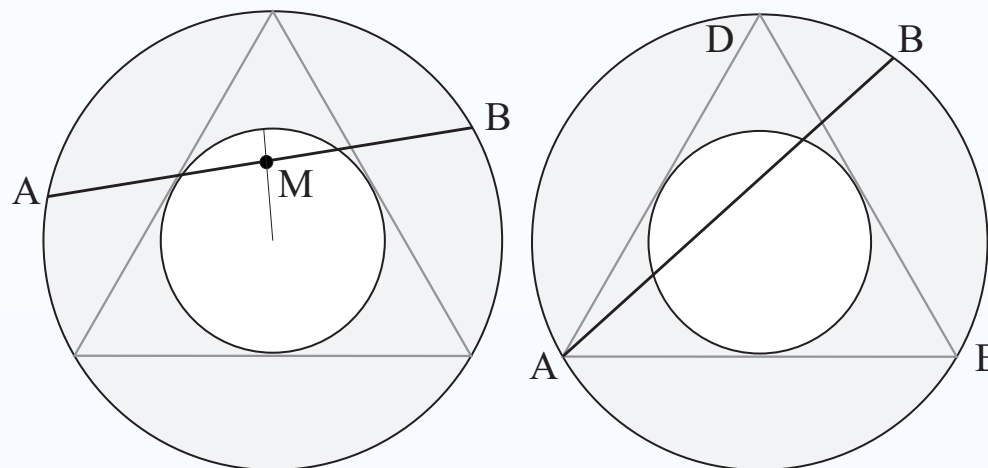
- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- Set Theory
- Properties of Axiomatic Probability
- Countable Spaces
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo



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.

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



Bertrand's Paradox

Aims and Objectives

Probability Theory

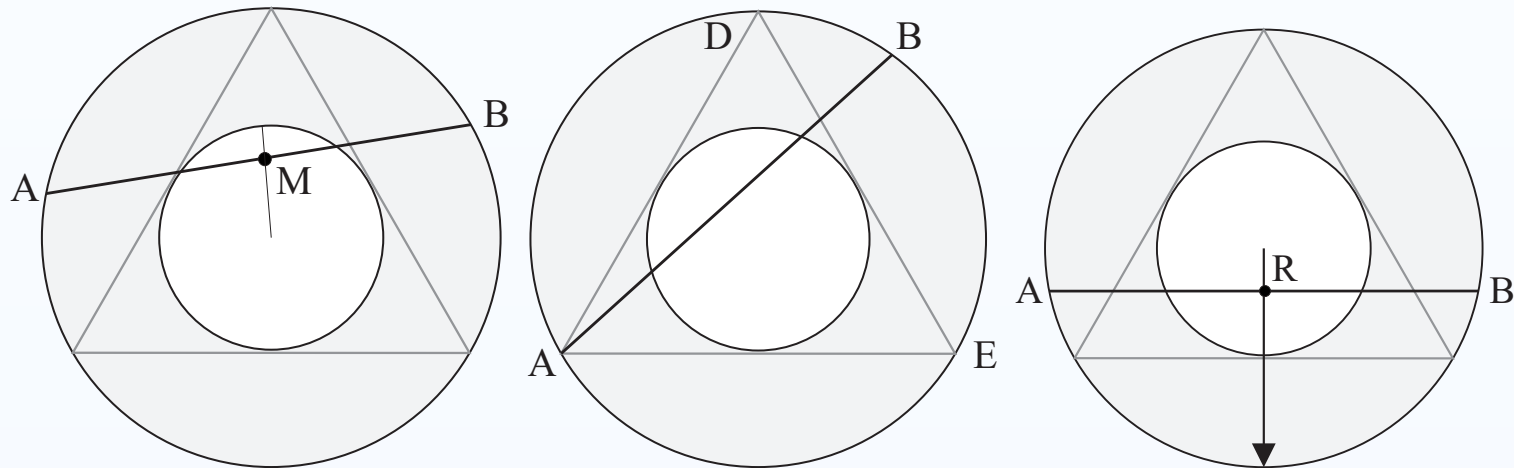
- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- Set Theory
- Properties of Axiomatic Probability
- Countable Spaces
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo



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?



Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- **Difficulties with the Classical Definition**
- Axiomatic Definition
- Set Theory
- Properties of Axiomatic Probability
- Countable Spaces
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

Using the Classical Definition

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

Example (Rolling two dice). Two dice are rolled; find the probability, p , that the sum of the numbers shown equals 7.

Consider three possibilities:

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



Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- **Difficulties with the Classical Definition**
- Axiomatic Definition
- Set Theory
- Properties of Axiomatic Probability
- Countable Spaces
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

Using the Classical Definition

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

Example (Rolling two dice). Two dice are rolled; find the probability, p , that the sum of the numbers shown equals 7.

Consider three possibilities:

1. The *possible outcomes* total 11 which are the sums $\{2, 3, \dots, 12\}$. Of these, only one (the sum 7) is favourable. Hence, $p = \frac{1}{11}$.
2. Therefore, to count all possible outcomes which are equally probable, it is necessary to could all pairs of numbers distinguishing between the first and second die. This will give the correct probability.



Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- **Difficulties with the Classical Definition**
- Axiomatic Definition
- Set Theory
- Properties of Axiomatic Probability
- Countable Spaces
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

Using the Classical Definition

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

Example (Rolling two dice). Two dice are rolled; find the probability, p , that the sum of the numbers shown equals 7.

Consider three possibilities:

1. The *possible outcomes* total 11 which are the sums $\{2, 3, \dots, 12\}$. Of these, only one (the sum 7) is favourable. Hence, $p = \frac{1}{11}$.
2. Therefore, to count all possible outcomes which are equally probable, it is necessary to could all pairs of numbers distinguishing between the first and second die. This will give the correct probability.



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!

Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- **Axiomatic Definition**
- Set Theory
- Properties of Axiomatic Probability
- Countable Spaces
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo



Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- **Set Theory**
- Properties of Axiomatic Probability
- Countable Spaces
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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.



Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- **Set Theory**
- Properties of Axiomatic Probability
- Countable Spaces
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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 ration is needed, such as length, or area. This leads to difficulties, as discussed in Bertrand's paradox.



Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- **Set Theory**
- Properties of Axiomatic Probability
- Countable Spaces
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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:

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

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

Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- **Set Theory**
- Properties of Axiomatic Probability
- Countable Spaces
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo



Set Theory

Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- **Set Theory**
- Properties of Axiomatic Probability
- Countable Spaces
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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

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



Set Theory

Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- Set Theory
- Properties of Axiomatic Probability
- Countable Spaces
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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

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

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

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

Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- Set Theory
- **Properties of Axiomatic Probability**
- Countable Spaces
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo



Set Theory

Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- Set Theory
- **Properties of Axiomatic Probability**
- Countable Spaces
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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

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

As an application of this, note that:

$$\begin{aligned} \overline{A \cup BC} &= \bar{A} \overline{BC} = \bar{A} (\bar{B} \cup \bar{C}) \\ &= (\bar{A} \bar{B}) \cup (\bar{A} \bar{C}) \\ &= \overline{A \cup B} \cup \overline{A \cup C} \end{aligned}$$

$$\Rightarrow \overline{A \cup BC} = (\bar{A} \cup \bar{B}) (\bar{A} \cup \bar{C})$$



Properties of Axiomatic Probability

Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- Set Theory
- Properties of Axiomatic Probability
- Countable Spaces
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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

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



Properties of Axiomatic Probability

Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- Set Theory
- Properties of Axiomatic Probability
- Countable Spaces
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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

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

Complements Since $A \cup \bar{A} = S$ and $A\bar{A} = \{\emptyset\}$, then

$$\Pr(A \cup \bar{A}) = \Pr(A) + \Pr(\bar{A}) = \Pr(S) = 1, \text{ such that:}$$

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



Properties of Axiomatic Probability

Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- Set Theory
- **Properties of Axiomatic Probability**
- Countable Spaces
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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

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

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

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

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

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



Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- Set Theory
- Properties of Axiomatic Probability
- Countable Spaces
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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

$$A \cup (\bar{A} B) = (A \cup \bar{A}) (A \cup B) = A \cup B$$

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

● Second, note that:

$$B = (A \cup \bar{A}) B = (A B) \cup (\bar{A} B) \quad \square$$

and that $(A B) \cap (\bar{A} B) = A \bar{A} B = \{\emptyset\} B = \{\emptyset\}$ and are therefore mutually exclusive events.



Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- Set Theory
- Properties of Axiomatic Probability
- Countable Spaces
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

Properties of Axiomatic Probability

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

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

$$\Pr(B) = \Pr((AB) \cup (\overline{A}B)) = \Pr(AB) + \Pr(\overline{A}B)$$

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

$$\Pr(A \cup B) - \Pr(B) = \Pr(A \cup (\overline{A}B)) - \Pr((AB) \cup (\overline{A}B)) = \Pr(A) - \Pr(AB) \quad \square$$



Properties of Axiomatic Probability

Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- Set Theory
- Properties of Axiomatic Probability
- **Countable Spaces**
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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



Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- Set Theory
- Properties of Axiomatic Probability
- **Countable Spaces**
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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

SOLUTION. Using the sum rule, that:

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

which is the case when the whole **sample space** is covered by the two events. 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 $\Pr(A \cap B) \leq \min\{\Pr(A), \Pr(B)\} = 1/3$.



Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- Set Theory
- Properties of Axiomatic Probability
- **Countable Spaces**
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

Countable Spaces

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

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



Countable Spaces

Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- Set Theory
- Properties of Axiomatic Probability
- **Countable Spaces**
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

Example (Cups and Saucers). SOLUTION. 📍 Lay the saucers in order, say as $RRWWBB$.



Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- Set Theory
- Properties of Axiomatic Probability
- Countable Spaces
- **The Real Line**
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

Countable Spaces

Example (Cups and Saucers). SOLUTION. 🍷 Lay the saucers in order, say as $RRWWBB$.

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



Countable Spaces

Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- Set Theory
- Properties of Axiomatic Probability
- Countable Spaces
- **The Real Line**
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

Example (Cups and Saucers). SOLUTION. 🍷 Lay the saucers in order, say as $RRWWBB$.

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

🍷 The arrangements in which cups never match their saucers are:

WWBBRR, WBRRWR, BWBRRW, BBRRWW

WBRRWR, BWRBRW

WBRRWR, BWRBWR

WBRRWR, BWBRRW





Countable Spaces

Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- Set Theory
- Properties of Axiomatic Probability
- Countable Spaces
- **The Real Line**
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

Example (Cups and Saucers). SOLUTION. 🍷 Lay the saucers in order, say as $RRWWBB$.

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

🍷 The arrangements in which cups never match their saucers are:

WWBBRR, WBRRWR, BWBRRW, BBRRWW

WBRRWR, BWRBRW

WBRRRW, BWRBWR

WBRRWR, BWBRRW



🍷 Hence, the required probability is $10/90 = 1/9$.



Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- Set Theory
- Properties of Axiomatic Probability
- Countable Spaces
- **The Real Line**
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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.



Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- Set Theory
- Properties of Axiomatic Probability
- Countable Spaces
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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.



Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- Set Theory
- Properties of Axiomatic Probability
- Countable Spaces
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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 $\{x \leq x_i\}$.



Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- Set Theory
- Properties of Axiomatic Probability
- Countable Spaces
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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.



Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- Set Theory
- Properties of Axiomatic Probability
- Countable Spaces
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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.



Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- Set Theory
- Properties of Axiomatic Probability
- Countable Spaces
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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:

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

provided that n is sufficiently large.

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



Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- Set Theory
- Properties of Axiomatic Probability
- Countable Spaces
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

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

Aims and Objectives

Probability Theory

- Introduction
- Classical Definition of Probability
- Bertrand's Paradox
- Using the Classical Definition
- Difficulties with the Classical Definition
- Axiomatic Definition
- Set Theory
- Properties of Axiomatic Probability
- Countable Spaces
- The Real Line
- Conditional Probability

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

Handout 3

Scalar Random Variables



Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

Abstract

- Deterministic signals interesting because their *signal values* are uniquely specified by a functional form.



Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- **Definition**
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

Abstract

- Deterministic signals interesting because their *signal values* are uniquely specified by a functional form.
- This precise description cannot be obtained for real-world signals.



Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

Abstract

- Deterministic signals interesting because their *signal values* are uniquely specified by a functional form.
- This precise description cannot be obtained for real-world signals.
- Moreover, it can be argued that real-world signals are inherently *stochastic* in nature.



Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

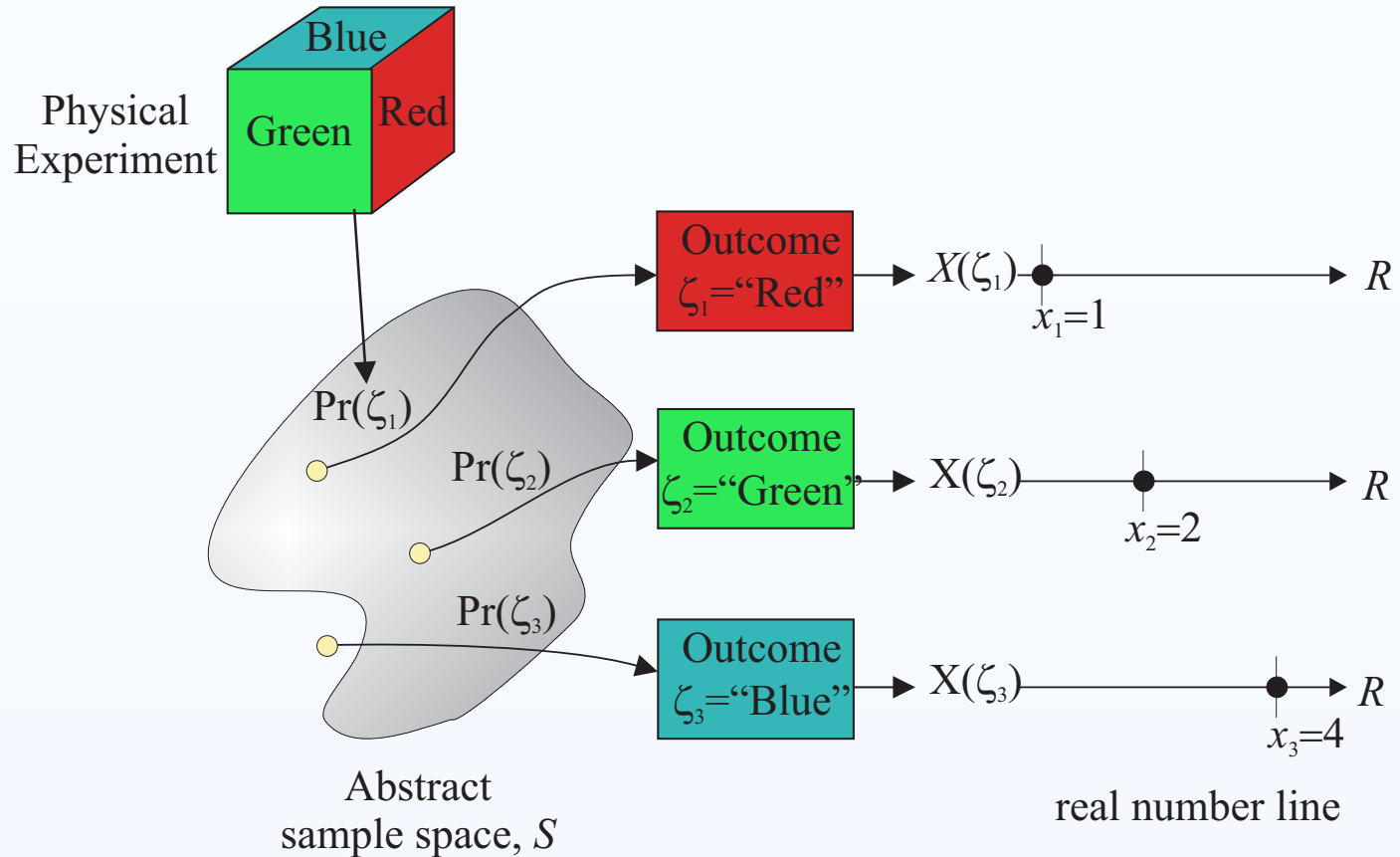
MonteCarlo

Abstract

- Deterministic signals interesting because their *signal values* are uniquely specified by a functional form.
- This precise description cannot be obtained for real-world signals.
- Moreover, it can be argued that real-world signals are inherently *stochastic* in nature.
- Although random signals evolve in time stochastically, their average properties are often deterministic, and thus can be specified by an explicit functional form.



Definition



A graphical representation of a random variable.

Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo



Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

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



Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

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:

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

✕



Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- **Density functions**
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

Distribution functions

- 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:

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



Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- **Properties: Distributions and Densities**
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

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:

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



Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- **Properties: Distributions and Densities**
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

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:

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

- It directly follows that:

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



Density functions

Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- **Properties: Distributions and Densities**
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

- 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:

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

- 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 : $p_k \triangleq \Pr(X(\zeta) = x_k)$.



Properties: Distributions and Densities

📌 Properties of cdf:

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

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

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

Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo



Properties: Distributions and Densities

Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

📍 Properties of cdf:

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

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

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

📍 Properties of pdfs:

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



Properties: Distributions and Densities

Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

📍 Properties of cdf:

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

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

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

📍 Properties of pdfs:

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

📍 Probability of arbitrary events:

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



Kolmogorov's Axioms

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

$$\begin{aligned}\Pr(x \leq x_1) + \Pr(x_1 < x \leq x_2) &= \Pr(x \leq x_2) \\ \int_{-\infty}^{x_1} p(v) dv + \Pr(x_1 < x \leq x_2) &= \int_{-\infty}^{x_2} p(v) dv \\ \Rightarrow \Pr(x_1 < x \leq x_2) &= \int_{x_1}^{x_2} p(v) dv\end{aligned}$$

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

Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo



Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- **Probability transformation rule**
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

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 has mean μ_X , but its variance does not exist.



Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

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 $f_Y(y)$?



Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

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 $f_Y(y)$?

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

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





Probability transformation rule

Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- **Probability transformation rule**
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

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 $f_Y(y)$?

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

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

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

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



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



Probability transformation rule

Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

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



Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- **Expectations**
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

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:

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





Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- **Expectations**
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

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



Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- **Expectations**
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

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:

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



Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

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:

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



Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

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

Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

📍 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.



Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

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.



Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

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 $\mu_X = a$.



Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

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:

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



Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

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:

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

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

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



Moments

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

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

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

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

Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo



Moments

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

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

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

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

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

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

Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo



Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

Characteristic Functions

The characteristic function of a rv $X(\zeta)$ is defined by the integral:

$$\Phi_X(\xi) \triangleq \mathbb{E} \left[e^{j\xi X(\zeta)} \right] = \int_{-\infty}^{\infty} f_X(x) e^{j\xi x} dx$$



Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

Characteristic Functions

The **characteristic function** of a rv $X(\zeta)$ is defined by the integral:

$$\Phi_X(\xi) \triangleq \mathbb{E} \left[e^{j\xi X(\zeta)} \right] = \int_{-\infty}^{\infty} f_X(x) e^{j\xi x} dx$$

When $j\xi$ is replaced by a complex variable s , the **moment generating function** is obtained, as defined by:

$$\bar{\Phi}_X(s) \triangleq \mathbb{E} \left[e^{sX(\zeta)} \right] = \int_{-\infty}^{\infty} f_X(x) e^{sx} dx$$



Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

Characteristic Functions

The **characteristic function** of a rv $X(\zeta)$ is defined by the integral:

$$\Phi_X(\xi) \triangleq \mathbb{E} \left[e^{j\xi X(\zeta)} \right] = \int_{-\infty}^{\infty} f_X(x) e^{j\xi x} dx$$

When $j\xi$ is replaced by a complex variable s , the **moment generating function** is obtained, as defined by:

$$\bar{\Phi}_X(s) \triangleq \mathbb{E} \left[e^{sX(\zeta)} \right] = \int_{-\infty}^{\infty} f_X(x) e^{sx} dx$$

Using a series expansion for $e^{sX(\zeta)}$ gives:

$$\bar{\Phi}_X(s) = \sum_{n=0}^{\infty} \frac{s^n}{n!} r_X^{(n)}$$

Thus, if all moments of $X(\zeta)$ are known upon inverse Laplace transformation, the pdf $f_X(x)$ can be determined.



Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

Characteristic Functions

$$\bar{\Phi}_X(s) = \sum_{n=0}^{\infty} \frac{s^n}{n!} r_X^{(n)}$$

Differentiating $\bar{\Phi}_X(s)$ m -times w. r. t. s , provides the m th-order moment of the RV $X(\zeta)$:

$$r_X^{(m)} = \left. \frac{d^m \bar{\Phi}_X(s)}{ds^m} \right|_{s=0} = (-j)^m \left. \frac{d^m \Phi_X(\xi)}{d\xi^m} \right|_{\xi=0}, \quad m \in \mathbb{Z}^+$$



Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

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 about its mean. It is defined as 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.

In otherwords, if the left side or *left tail* of the distribution is more pronounced than the *right tail*, the function is said to have negative skewness (and leans to the left). If the reverse is true, it has positive skewness (and leans to the right). If the two are equal, it has zero skewness.



Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

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 about its mean. It is defined as 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:

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

In otherwords, if the left side or *left tail* of the distribution is



Higher-order statistics

Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

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

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.



Aims and Objectives

Probability Theory

Scalar Random Variables

- Abstract
- Definition
- Distribution functions
- Density functions
- Properties: Distributions and Densities
- Kolmogorov's Axioms
- Common Continuous RVs
- Probability transformation rule
- Expectations
- Properties of expectation operator
- Moments
- Characteristic Functions
- Higher-order statistics

Multiple Random Variables

Estimation Theory

MonteCarlo

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



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

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.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

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*.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

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.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

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.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- **Distribution and Density Functions**
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

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:

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



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- **Distribution and Density Functions**
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

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 .



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- **Distribution and Density Functions**
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

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:

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

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- **Distribution and Density Functions**
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



Distribution and Density Functions

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

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

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

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

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- **Distribution and Density Functions**
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- **Distribution and Density Functions**
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Distribution and Density Functions

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

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

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

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

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

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- **Distribution and Density Functions**
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

📍 Properties of joint-cdf:

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

$F_{\mathbf{X}}(\mathbf{x})$ 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

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- **Distribution and Density Functions**
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

📍 Properties of joint-cdf:

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

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

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

📍 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

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- **Distribution and Density Functions**
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

📍 Properties of joint-cdf:

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

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

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

📍 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

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- **Distribution and Density Functions**
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

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, $F_{\mathbf{Z}}(\mathbf{z})$.



Distribution and Density Functions

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

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:

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





Distribution and Density Functions

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

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:

$$\begin{aligned} \int_{-\infty}^{\infty} f_{\mathbf{Z}}(\mathbf{z}) d\mathbf{z} &= \int_0^1 \int_0^1 \frac{1}{2}(x + 3y) dx dy = \int_0^1 \frac{1}{2} \left[\frac{1}{2}x^2 + 3xy \right]_0^1 dy \\ &= \int_0^1 \frac{1}{4} + \frac{3}{2}y dy = \left[\frac{y}{4} + \frac{3y^2}{4} \right]_0^1 = \frac{1}{4} + \frac{3}{4} = 1 \end{aligned}$$





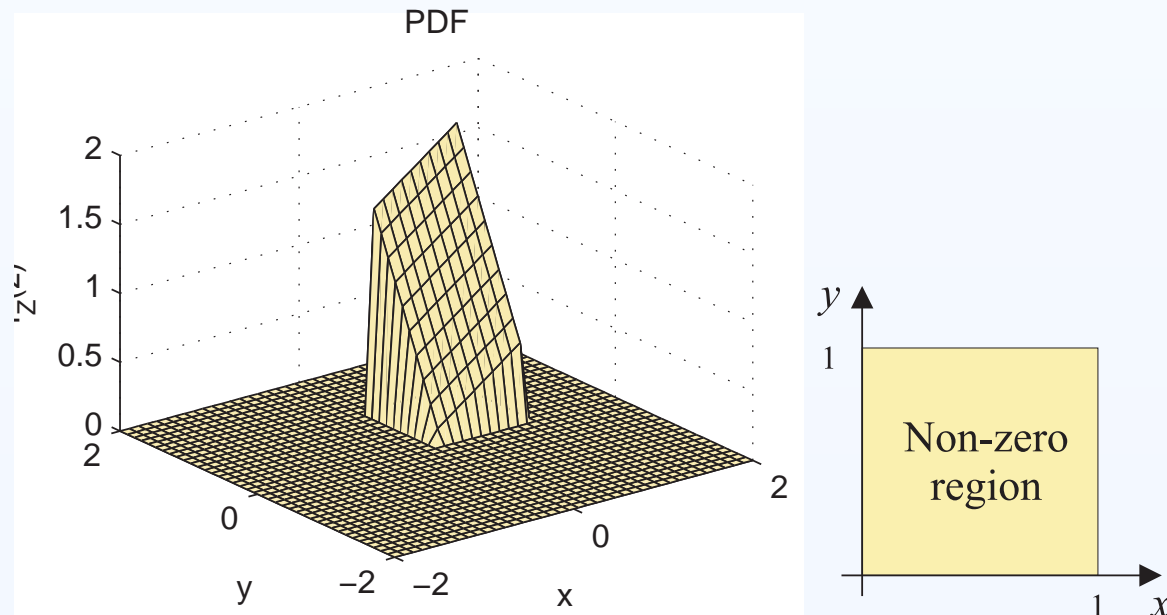
Distribution and Density Functions

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 pdf is shown here:



Region of support for pdf.

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



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

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- **Distribution and Density Functions**
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



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

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

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



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



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

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

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

□

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



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

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

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

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

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



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:

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

□

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- **Marginal Density Function**
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



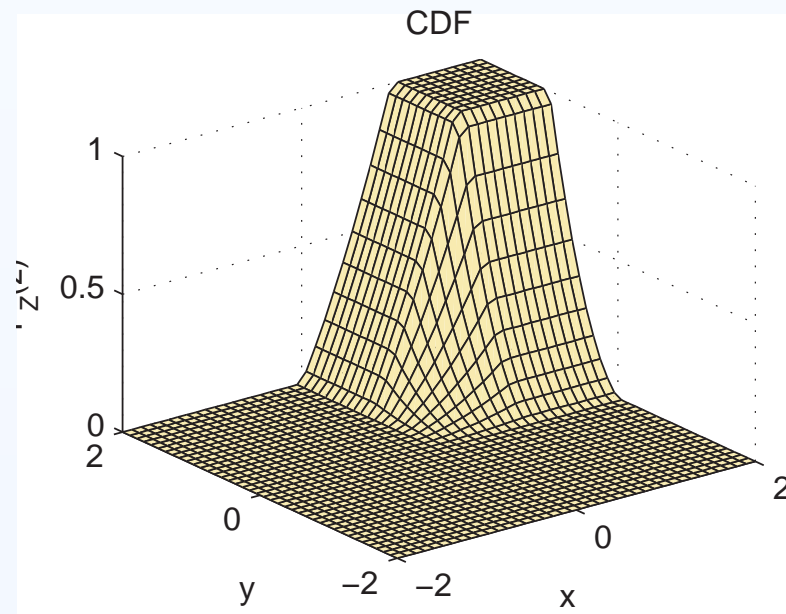
Distribution and Density Functions

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.

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- **Marginal Density Function**
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Marginal Density Function

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



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- **Marginal Density Function**
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

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 $\mathbf{X}(\zeta)$,

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



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- **Marginal Density Function**
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

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:

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

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- **Marginal Density Function**
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



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 :

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

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



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

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- **Marginal Density Function**
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- **Marginal Density Function**
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

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 $F_Y(y)$.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- **Marginal Density Function**
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

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 $F_Y(y)$.

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:

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



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- **Marginal Density Function**
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



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:

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

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



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

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

□

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



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

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

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

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



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

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

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



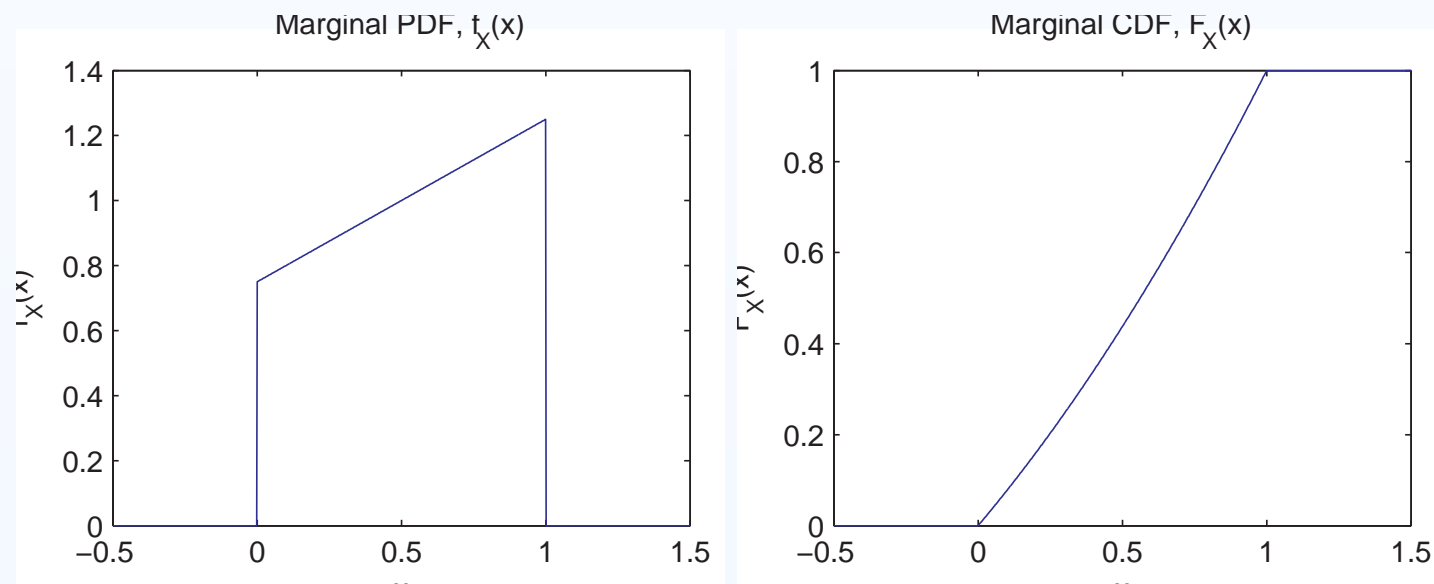


Marginal Density Function

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, $X(\zeta)$.

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's Rule
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

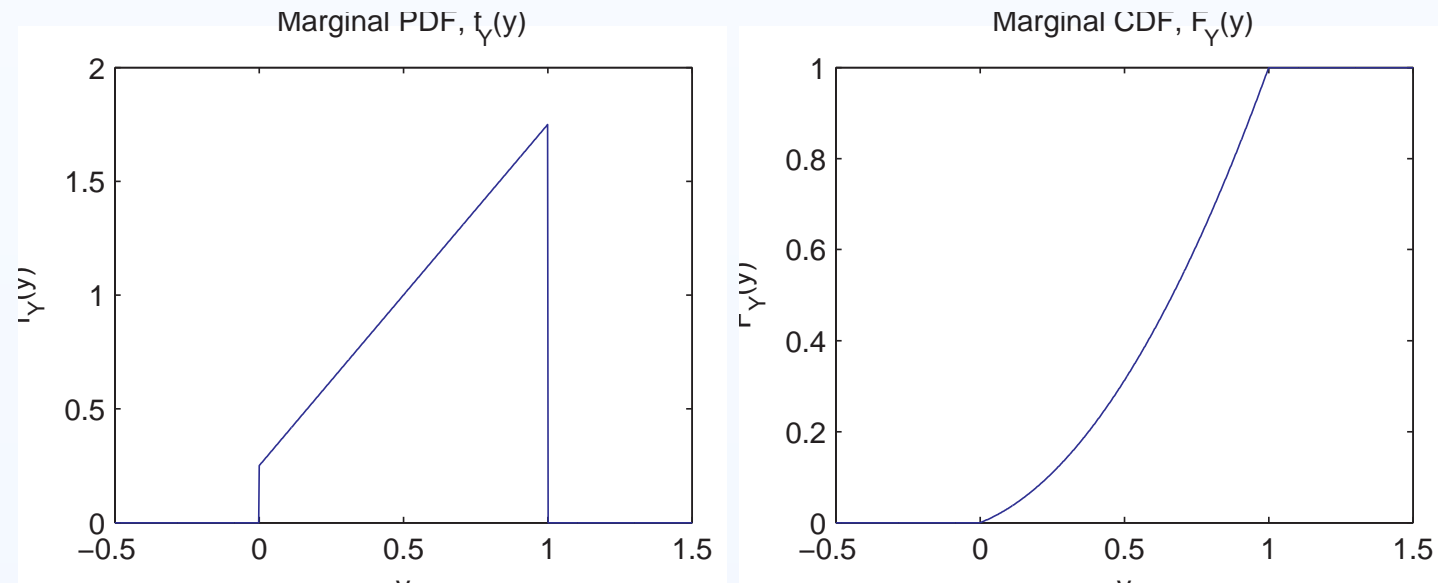


Marginal Density Function

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, $Y(\zeta)$.

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



Independence

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

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

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



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

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

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

This then implies that

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

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



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

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 $\mathbf{X}(\zeta)$.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- **Conditionals and Bayes's**
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

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:

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



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- **Statistical Description**
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

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:

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

If the random vectors $\mathbf{X}(\zeta)$ and $\mathbf{Y}(\zeta)$ are independent, then the conditional pdf must be identical to the unconditional pdf:

$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x}) = f_{\mathbf{Y}}(\mathbf{y})$. Hence, it follows that:

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



Conditionals and Bayes's

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- **Statistical Description**
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Since

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

it follows

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



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- **Statistical Description**
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Conditionals and Bayes's

Since

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

it follows

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

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

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

then it follows

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



Statistical Description

Statistical averages are more manageable, but less of a complete description of random vectors.

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- **Statistical Description**
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- **Statistical Description**
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

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.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- **Statistical Description**
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

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.

However, it is important to understand that multiple RVs leads to the notion of measuring their interaction or dependence. This concept is useful in abstract, but also when dealing with stochastic processes or time-series.



Statistical Description

Mean vector The mean vector is the first-moment of the random vector, and is given by:

$$\mu_{\mathbf{X}} = \mathbb{E} [\mathbf{X} (\zeta)] = \begin{bmatrix} \mathbb{E} [X_1(\zeta)] \\ \vdots \\ \mathbb{E} [X_N(\zeta)] \end{bmatrix} = \begin{bmatrix} \mu_{X_1} \\ \vdots \\ \mu_{X_N} \end{bmatrix}$$

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- **Statistical Description**
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



Statistical Description

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- **Statistical Description**
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Mean vector The mean vector is the first-moment of the random vector, and is given by:

$$\mu_{\mathbf{X}} = \mathbb{E} [\mathbf{X} (\zeta)] = \begin{bmatrix} \mathbb{E} [X_1(\zeta)] \\ \vdots \\ \mathbb{E} [X_N(\zeta)] \end{bmatrix} = \begin{bmatrix} \mu_{X_1} \\ \vdots \\ \mu_{X_N} \end{bmatrix}$$

Correlation Matrix The second-order moments of the random vector describe the spread of the distribution. The **autocorrelation matrix** is defined by:

$$\mathbf{R}_{\mathbf{X}} \triangleq \mathbb{E} [\mathbf{X} (\zeta) \mathbf{X}^H (\zeta)] = \begin{bmatrix} r_{X_1 X_1} & \cdots & r_{X_1 X_N} \\ \vdots & \ddots & \vdots \\ r_{X_N X_1} & \cdots & r_{X_N X_N} \end{bmatrix}$$



Statistical Description

Correlation Matrix The diagonal terms

$$r_{X_i X_i} \triangleq \mathbb{E} \left[|X_i(\zeta)|^2 \right], \quad i \in \{1, \dots, N\}$$

are the second-order moments of each of the RVs, $X_i(\zeta)$.

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- **Statistical Description**
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



Statistical Description

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- **Statistical Description**
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Correlation Matrix The diagonal terms

$$r_{X_i X_i} \triangleq \mathbb{E} \left[|X_i(\zeta)|^2 \right], \quad i \in \{1, \dots, N\}$$

are the second-order moments of each of the RVs, $X_i(\zeta)$.

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**, or statistical similarity between the RVs $X_i(\zeta)$ and $X_j(\zeta)$.



Statistical Description

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- **Statistical Description**
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Correlation Matrix The diagonal terms

$$r_{X_i X_i} \triangleq \mathbb{E} \left[|X_i(\zeta)|^2 \right], \quad i \in \{1, \dots, N\}$$

are the second-order moments of each of the RVs, $X_i(\zeta)$.

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**, or statistical similarity between the RVs $X_i(\zeta)$ and $X_j(\zeta)$.

If the $X_i(\zeta)$ and $X_j(\zeta)$ are **orthogonal** then their **correlation is zero**:

$$r_{X_i X_j} = \mathbb{E} \left[X_i(\zeta) X_j^*(\zeta) \right] = 0, \quad i \neq j$$



Statistical Description

Covariance Matrix The autocovariance matrix is defined by:

$$\Gamma_{\mathbf{X}} \triangleq \mathbb{E} \left[(\mathbf{X}(\zeta) - \boldsymbol{\mu}_{\mathbf{X}}) (\mathbf{X}(\zeta) - \boldsymbol{\mu}_{\mathbf{X}})^H \right] = \begin{bmatrix} \gamma_{X_1 X_1} & \cdots & \gamma_{X_1 X_N} \\ \vdots & \ddots & \dots \\ \gamma_{X_N X_1} & \cdots & \gamma_{X_N X_N} \end{bmatrix}$$

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- **Statistical Description**
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



Statistical Description

Covariance Matrix The autocovariance matrix is defined by:

$$\mathbf{\Gamma}_{\mathbf{X}} \triangleq \mathbb{E} \left[(\mathbf{X}(\zeta) - \boldsymbol{\mu}_{\mathbf{X}}) (\mathbf{X}(\zeta) - \boldsymbol{\mu}_{\mathbf{X}})^H \right] = \begin{bmatrix} \gamma_{X_1 X_1} & \cdots & \gamma_{X_1 X_N} \\ \vdots & \ddots & \dots \\ \gamma_{X_N X_1} & \cdots & \gamma_{X_N X_N} \end{bmatrix}$$

The diagonal terms

$$\gamma_{X_i X_i} \triangleq \sigma_{X_i}^2 = \mathbb{E} \left[|X_i(\zeta) - \mu_{X_i}|^2 \right], \quad i \in \{1, \dots, N\}$$

are the **variances** of each of the RVs, $X_i(\zeta)$.

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- **Statistical Description**
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



Statistical Description

Covariance Matrix The off-diagonal terms

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- **Statistical Description**
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

$$\begin{aligned}\gamma_{X_i X_j} &\triangleq \mathbb{E} \left[(X_i(\zeta) - \mu_{X_i}) (X_j(\zeta) - \mu_{X_j})^* \right] \\ &= r_{X_i X_j} - \mu_{X_i} \mu_{X_j}^* = \gamma_{X_j X_i}^*, \quad i \neq j\end{aligned}$$

measure the **covariance** $X_i(\zeta)$ and $X_j(\zeta)$.



Statistical Description

Covariance Matrix The off-diagonal terms

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- **Statistical Description**
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

$$\begin{aligned}\gamma_{X_i X_j} &\triangleq \mathbb{E} \left[(X_i(\zeta) - \mu_{X_i}) (X_j(\zeta) - \mu_{X_j})^* \right] \\ &= r_{X_i X_j} - \mu_{X_i} \mu_{X_j}^* = \gamma_{X_j X_i}^*, \quad i \neq j\end{aligned}$$

measure the **covariance** $X_i(\zeta)$ and $X_j(\zeta)$.

It should also be noticed that the **covariance** and **correlation** matrices are positive semidefinite; that is, they satisfy the relations:

$$\mathbf{a}^H \mathbf{R}_X \mathbf{a} \geq 0$$

$$\mathbf{a}^H \mathbf{\Gamma}_X \mathbf{a} \geq 0$$

for any complex vector \mathbf{a} .



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- **Statistical Description**
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Statistical Description

The autocorrelation and autocovariance matrices are related, and it can easily be seen that:

$$\Gamma_{\mathbf{X}} \triangleq \mathbb{E} \left[[\mathbf{X}(\zeta) - \boldsymbol{\mu}_{\mathbf{X}}] [\mathbf{X}(\zeta) - \boldsymbol{\mu}_{\mathbf{X}}]^H \right] = \mathbf{R}_{\mathbf{X}} - \boldsymbol{\mu}_{\mathbf{X}} \boldsymbol{\mu}_{\mathbf{X}}^H$$

In fact, if $\boldsymbol{\mu}_{\mathbf{X}} = 0$, then $\Gamma_{\mathbf{X}} = \mathbf{R}_{\mathbf{X}}$.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- **Statistical Description**
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Statistical Description

The autocorrelation and autocovariance matrices are related, and it can easily be seen that:

$$\Gamma_{\mathbf{X}} \triangleq \mathbb{E} \left[[\mathbf{X}(\zeta) - \boldsymbol{\mu}_{\mathbf{X}}] [\mathbf{X}(\zeta) - \boldsymbol{\mu}_{\mathbf{X}}]^H \right] = \mathbf{R}_{\mathbf{X}} - \boldsymbol{\mu}_{\mathbf{X}} \boldsymbol{\mu}_{\mathbf{X}}^H$$

In fact, if $\boldsymbol{\mu}_{\mathbf{X}} = \mathbf{0}$, then $\Gamma_{\mathbf{X}} = \mathbf{R}_{\mathbf{X}}$.

If the random variables $X_i(\zeta)$ and $X_j(\zeta)$ are **independent**, then they are also **uncorrelated** since:

$$\begin{aligned} r_{X_i X_j} &= \mathbb{E} [X_i(\zeta) X_j(\zeta)^*] = \mathbb{E} [X_i(\zeta)] \mathbb{E} [X_j^*(\zeta)] \\ &= \mu_{X_i} \mu_{X_j}^* \Rightarrow \gamma_{X_i X_j} = 0 \end{aligned}$$



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- **Statistical Description**
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Statistical Description

The autocorrelation and autocovariance matrices are related, and it can easily be seen that:

$$\Gamma_{\mathbf{X}} \triangleq \mathbb{E} \left[[\mathbf{X}(\zeta) - \boldsymbol{\mu}_{\mathbf{X}}] [\mathbf{X}(\zeta) - \boldsymbol{\mu}_{\mathbf{X}}]^H \right] = \mathbf{R}_{\mathbf{X}} - \boldsymbol{\mu}_{\mathbf{X}} \boldsymbol{\mu}_{\mathbf{X}}^H$$

In fact, if $\boldsymbol{\mu}_{\mathbf{X}} = 0$, then $\Gamma_{\mathbf{X}} = \mathbf{R}_{\mathbf{X}}$.

If the random variables $X_i(\zeta)$ and $X_j(\zeta)$ are **independent**, then they are also **uncorrelated** since:

$$\begin{aligned} r_{X_i X_j} &= \mathbb{E} [X_i(\zeta) X_j(\zeta)^*] = \mathbb{E} [X_i(\zeta)] \mathbb{E} [X_j^*(\zeta)] \\ &= \mu_{X_i} \mu_{X_j}^* \Rightarrow \gamma_{X_i X_j} = 0 \end{aligned}$$

Note, however, that uncorrelatedness does not imply independence, unless the RVs are jointly-Gaussian.



Statistical Description

Cross-correlation is defined as

$$\mathbf{R}_{\mathbf{X}\mathbf{Y}} \triangleq \mathbb{E} [\mathbf{X}(\zeta) \mathbf{Y}^H(\zeta)] = \begin{bmatrix} \mathbb{E} [X_1(\zeta) Y_1^*(\zeta)] & \cdots & \mathbb{E} [X_1(\zeta) Y_M^*(\zeta)] \\ \vdots & \ddots & \vdots \\ \mathbb{E} [X_N(\zeta) Y_1^*(\zeta)] & \cdots & \mathbb{E} [X_N(\zeta) Y_M^*(\zeta)] \end{bmatrix}$$

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



Statistical Description

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- **Probability Transformation Rule**
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Cross-correlation is defined as

$$\mathbf{R}_{\mathbf{XY}} \triangleq \mathbb{E} [\mathbf{X}(\zeta) \mathbf{Y}^H(\zeta)] = \begin{bmatrix} \mathbb{E} [X_1(\zeta) Y_1^*(\zeta)] & \cdots & \mathbb{E} [X_1(\zeta) Y_M^*(\zeta)] \\ \vdots & \ddots & \vdots \\ \mathbb{E} [X_N(\zeta) Y_1^*(\zeta)] & \cdots & \mathbb{E} [X_N(\zeta) Y_M^*(\zeta)] \end{bmatrix}$$

Cross-covariance is defined as

$$\begin{aligned} \mathbf{\Gamma}_{\mathbf{XY}} &\triangleq \mathbb{E} \left[\{\mathbf{X}(\zeta) - \boldsymbol{\mu}_{\mathbf{X}}\} \{\mathbf{Y}(\zeta) - \boldsymbol{\mu}_{\mathbf{Y}}\}^H \right] \\ &= \mathbf{R}_{\mathbf{XY}} - \boldsymbol{\mu}_{\mathbf{X}} \boldsymbol{\mu}_{\mathbf{Y}}^H \end{aligned}$$



Statistical Description

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Cross-correlation is defined as

$$\mathbf{R}_{\mathbf{X}\mathbf{Y}} \triangleq \mathbb{E} [\mathbf{X}(\zeta) \mathbf{Y}^H(\zeta)] = \begin{bmatrix} \mathbb{E} [X_1(\zeta) Y_1^*(\zeta)] & \cdots & \mathbb{E} [X_1(\zeta) Y_M^*(\zeta)] \\ \vdots & \ddots & \vdots \\ \mathbb{E} [X_N(\zeta) Y_1^*(\zeta)] & \cdots & \mathbb{E} [X_N(\zeta) Y_M^*(\zeta)] \end{bmatrix}$$

Cross-covariance is defined as

$$\begin{aligned} \mathbf{\Gamma}_{\mathbf{X}\mathbf{Y}} &\triangleq \mathbb{E} \left[\{\mathbf{X}(\zeta) - \boldsymbol{\mu}_{\mathbf{X}}\} \{\mathbf{Y}(\zeta) - \boldsymbol{\mu}_{\mathbf{Y}}\}^H \right] \\ &= \mathbf{R}_{\mathbf{X}\mathbf{Y}} - \boldsymbol{\mu}_{\mathbf{X}} \boldsymbol{\mu}_{\mathbf{Y}}^H \end{aligned}$$

🔴 Uncorrelated if $\mathbf{\Gamma}_{\mathbf{X}\mathbf{Y}} = \mathbf{0} \Rightarrow \mathbf{R}_{\mathbf{X}\mathbf{Y}} = \boldsymbol{\mu}_{\mathbf{X}} \boldsymbol{\mu}_{\mathbf{Y}}^H$.

🔴 Orthogonal if $\mathbf{R}_{\mathbf{X}\mathbf{Y}} = \mathbf{0}$.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- **Polar Transformation**
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

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:

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





Probability Transformation Rule

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- **Linear Transformations**
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

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:

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

Assuming M -real vector-roots of the equation $\mathbf{y} = \mathbf{g}(\mathbf{x})$ by $\{\mathbf{x}_m, m \in \mathcal{M}\}$,

$$\mathbf{y} = \mathbf{g}(\mathbf{x}_1) = \dots = \mathbf{g}(\mathbf{x}_M)$$

then the joint-pdf of $\mathbf{Y}(\zeta)$ in terms of (i. t. o.) the joint-pdf of $\mathbf{X}(\zeta)$ is:

$$f_{\mathbf{Y}}(\mathbf{y}) = \sum_{m=1}^M \frac{f_{\mathbf{X}}(\mathbf{x}_m)}{|J(\mathbf{x}_m)|}$$



The Jacobian is defined in the notes, but is the usual definition!



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

$$r(\zeta) = \sqrt{X^2(\zeta) + Y^2(\zeta)}$$

$$\theta(\zeta) = \arctan \frac{Y(\zeta)}{X(\zeta)}$$

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



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:

$$f_{R,\Theta}(r, \theta) = r f_{XY}(r \cos \theta, r \sin \theta)$$

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- **Linear Transformations**
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Linear Transformations

Since linear systems represent such an important class of signal processing systems, it is important to consider **linear transformations** of random vectors.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Linear Transformations

Since linear systems represent such an important class of signal processing systems, it is important to consider **linear transformations** of random vectors.

Thus, consider a random vector $\mathbf{Y}(\zeta)$ defined by a linear transformation of the random vector $\mathbf{X}(\zeta)$ through the matrix \mathbf{A} :

$$\mathbf{Y}(\zeta) = \mathbf{A} \mathbf{X}(\zeta)$$



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- **Linear Transformations**
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Linear Transformations

Since linear systems represent such an important class of signal processing systems, it is important to consider **linear transformations** of random vectors.

Thus, consider a random vector $\mathbf{Y}(\zeta)$ defined by a linear transformation of the random vector $\mathbf{X}(\zeta)$ through the matrix \mathbf{A} :

$$\mathbf{Y}(\zeta) = \mathbf{A} \mathbf{X}(\zeta)$$

The Jacobian of a nonsingular linear transformation defined by a matrix \mathbf{A} is simply the absolute value of the determinant of \mathbf{A} . Thus, assuming $\mathbf{X}(\zeta)$, $\mathbf{Y}(\zeta)$, and \mathbf{A} are all real, then:

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{f_{\mathbf{X}}(\mathbf{A}^{-1}\mathbf{y})}{|\det \mathbf{A}|}$$



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Linear Transformations

In general, determining $f_{\mathbf{Y}}(\mathbf{y})$ is a laborious exercise, except in the case of Gaussian random vectors.

In practice, however, the knowledge of $\mu_{\mathbf{Y}}$, $\Gamma_{\mathbf{Y}}$, $\Gamma_{\mathbf{X}\mathbf{Y}}$ or $\Gamma_{\mathbf{Y}\mathbf{X}}$ is sufficient information for many algorithms.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Linear Transformations

In general, determining $f_Y(\mathbf{y})$ is a laborious exercise, except in the case of Gaussian random vectors.

In practice, however, the knowledge of μ_Y , Γ_Y , Γ_{XY} or Γ_{YX} is sufficient information for many algorithms.

Taking expectations of both sides of Equation ??, the following relations are found:

Mean vector:

$$\mu_Y = \mathbb{E}[\mathbf{A} \mathbf{X}(\zeta)] = \mathbf{A} \mu_X$$



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Linear Transformations

In general, determining $f_{\mathbf{Y}}(\mathbf{y})$ is a laborious exercise, except in the case of Gaussian random vectors.

In practice, however, the knowledge of $\mu_{\mathbf{Y}}$, $\Gamma_{\mathbf{Y}}$, $\Gamma_{\mathbf{X}\mathbf{Y}}$ or $\Gamma_{\mathbf{Y}\mathbf{X}}$ is sufficient information for many algorithms.

Taking expectations of both sides of Equation ??, the following relations are found:

Mean vector:

$$\mu_{\mathbf{Y}} = \mathbb{E} [\mathbf{A} \mathbf{X}(\zeta)] = \mathbf{A} \mu_{\mathbf{X}}$$

Autocorrelation matrix:

$$\begin{aligned} \mathbf{R}_{\mathbf{Y}} &= \mathbb{E} [\mathbf{Y}(\zeta) \mathbf{Y}^H(\zeta)] = \mathbb{E} [\mathbf{A} \mathbf{X}(\zeta) \mathbf{X}^H(\zeta) \mathbf{A}^H] \\ &= \mathbf{A} \mathbb{E} [\mathbf{X}(\zeta) \mathbf{X}^H(\zeta)] \mathbf{A}^H = \mathbf{A} \mathbf{R}_{\mathbf{X}} \mathbf{A}^H \end{aligned}$$



Linear Transformations

Autocovariance matrix:

$$\Gamma_Y = A\Gamma_X A^H$$

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- **Auxiliary Variables**
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Linear Transformations

Autocovariance matrix:

$$\Gamma_Y = A\Gamma_X A^H$$

Cross-correlation matrix:

$$\begin{aligned} \mathbf{R}_{\mathbf{X}\mathbf{Y}} &= \mathbb{E} [\mathbf{X}(\zeta) \mathbf{Y}^H(\zeta)] = \mathbb{E} [\mathbf{X}(\zeta) \mathbf{X}^H(\zeta) \mathbf{A}^H] \\ &= \mathbb{E} [\mathbf{X}(\zeta) \mathbf{X}^H(\zeta)] \mathbf{A}^H = \mathbf{R}_X \mathbf{A}^H \end{aligned}$$

and hence $\mathbf{R}_{\mathbf{Y}\mathbf{X}} = \mathbf{A}\mathbf{R}_X$.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Linear Transformations

Autocovariance matrix:

$$\Gamma_{\mathbf{Y}} = \mathbf{A}\Gamma_{\mathbf{X}}\mathbf{A}^H$$

Cross-correlation matrix:

$$\begin{aligned}\mathbf{R}_{\mathbf{X}\mathbf{Y}} &= \mathbb{E} [\mathbf{X}(\zeta) \mathbf{Y}^H(\zeta)] = \mathbb{E} [\mathbf{X}(\zeta) \mathbf{X}^H(\zeta) \mathbf{A}^H] \\ &= \mathbb{E} [\mathbf{X}(\zeta) \mathbf{X}^H(\zeta)] \mathbf{A}^H = \mathbf{R}_{\mathbf{X}} \mathbf{A}^H\end{aligned}$$

and hence $\mathbf{R}_{\mathbf{Y}\mathbf{X}} = \mathbf{A}\mathbf{R}_{\mathbf{X}}$.

Cross-covariance matrices:

$$\Gamma_{\mathbf{X}\mathbf{Y}} = \Gamma_{\mathbf{X}} \mathbf{A}^H \quad \text{and} \quad \Gamma_{\mathbf{Y}\mathbf{X}} = \mathbf{A} \Gamma_{\mathbf{X}}$$

These results will be used to show what happens to a Gaussian random vector under a linear transformation .



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- **Auxiliary Variables**
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Auxiliary Variables

The density of a RV that is *one* function $Z(\zeta) = g(X(\zeta), Y(\zeta))$ of two RVs can be determined by choosing a **auxiliary variable**.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- **Multivariate Gaussian Density Function**
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Auxiliary Variables

The density of a RV that is *one* function $Z(\zeta) = g(X(\zeta), Y(\zeta))$ of two RVs can be determined by choosing a **auxiliary variable**.

$$f_Z(z) = \int_{\mathbb{R}} f_{WZ}(w, z) dw = \sum_{m=1}^M \int_{\mathbb{R}} \frac{f_{\mathbf{XY}}(x_m, y_m)}{|J(x_m, y_m)|} dw$$



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- **Multivariate Gaussian Density Function**
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Auxiliary Variables

The density of a RV that is *one* function $Z(\zeta) = g(X(\zeta), Y(\zeta))$ of two RVs can be determined by choosing a **auxiliary variable**.

$$f_Z(z) = \int_{\mathbb{R}} f_{WZ}(w, z) dw = \sum_{m=1}^M \int_{\mathbb{R}} \frac{f_{\mathbf{XY}}(x_m, y_m)}{|J(x_m, y_m)|} dw$$

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

The density of a RV that is *one* function $Z(\zeta) = g(X(\zeta), Y(\zeta))$ of two RVs can be determined by choosing a **auxiliary variable**.

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- **Multivariate Gaussian Density Function**
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

$$f_Z(z) = \int_{\mathbb{R}} f_{WZ}(w, z) dw = \sum_{m=1}^M \int_{\mathbb{R}} \frac{f_{\mathbf{XY}}(x_m, y_m)}{|J(x_m, y_m)|} dw$$

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 $x = \frac{z-bw}{a}, y = w$.

Thus:

$$f_Z(z) = \frac{1}{|a|} \int_{\mathbb{R}} f_{XY}\left(\frac{z-bw}{a}, w\right) dw$$

□



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- **Multivariate Gaussian Density Function**
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

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 or Gaussian density function.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- **Multivariate Gaussian Density Function**
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

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 or Gaussian density function.

For a *real* random vector, this density function has the form:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{N}{2}} |\mathbf{\Gamma}_{\mathbf{X}}|^{\frac{1}{2}}} \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}})^T \mathbf{\Gamma}_{\mathbf{X}}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}}) \right]$$

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:

$$f_{\mathbf{X}}(\mathbf{x}) = \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_{\mathbf{X}}, \mathbf{\Gamma}_{\mathbf{X}})$$



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Multivariate Gaussian Density Function

The normal distribution is a useful model of a random vector because of its many important properties.

1. $f_{\mathbf{X}}(\mathbf{x}) = \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\Gamma}_{\mathbf{X}})$ 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.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- **Characteristic Functions**
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Multivariate Gaussian Density Function

The normal distribution is a useful model of a random vector because of its many important properties.

1. $f_{\mathbf{X}}(\mathbf{x}) = \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\Gamma}_{\mathbf{X}})$ 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.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- **Characteristic Functions**
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Multivariate Gaussian Density Function

The normal distribution is a useful model of a random vector because of its many important properties.

1. $f_{\mathbf{X}}(\mathbf{x}) = \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\Gamma}_{\mathbf{X}})$ 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.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- **Characteristic Functions**
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Characteristic Functions

The **characteristic function** and **moment generating function** for a scalar random variable can be extended to deal with random vectors. Essentially, these are defined as the multi-dimensional Fourier transform of the joint-pdf.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- **Characteristic Functions**
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Characteristic Functions

The **characteristic function** and **moment generating function** for a scalar random variable can be extended to deal with random vectors. Essentially, these are defined as the multi-dimensional Fourier transform of the joint-pdf.

Hence, the characteristic function is:

$$\Phi_{\mathbf{X}}(\boldsymbol{\xi}) \triangleq \mathbb{E} \left[e^{j\boldsymbol{\xi}^T \mathbf{X}(\zeta)} \right] = \int_{-\infty}^{\infty} f_{\mathbf{X}}(\mathbf{x}) e^{j\boldsymbol{\xi}^T \mathbf{x}} d\mathbf{x}$$



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- **Characteristic Functions**
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Characteristic Functions

The **characteristic function** and **moment generating function** for a scalar random variable can be extended to deal with random vectors. Essentially, these are defined as the multi-dimensional Fourier transform of the joint-pdf.

Hence, the characteristic function is:

$$\Phi_{\mathbf{X}}(\boldsymbol{\xi}) \triangleq \mathbb{E} \left[e^{j\boldsymbol{\xi}^T \mathbf{X}(\zeta)} \right] = \int_{-\infty}^{\infty} f_{\mathbf{X}}(\mathbf{x}) e^{j\boldsymbol{\xi}^T \mathbf{x}} d\mathbf{x}$$

Similarly, the moment generating function is given by:

$$\bar{\Phi}_{\mathbf{X}}(\mathbf{s}) \triangleq \mathbb{E} \left[e^{\mathbf{s}^T \mathbf{X}(\zeta)} \right] = \int_{-\infty}^{\infty} f_{\mathbf{X}}(\mathbf{x}) e^{\mathbf{s}^T \mathbf{x}} d\mathbf{x}$$



Characteristic Functions

The characteristic function for a Gaussian random vector is

$$\begin{aligned}\Phi_{\mathbf{X}}(\boldsymbol{\xi}) &= \int_{-\infty}^{\infty} f_{\mathbf{X}}(\mathbf{x}) e^{j\boldsymbol{\xi}^T \mathbf{x}} d\mathbf{x} \\ &= \frac{1}{(2\pi)^{\frac{N}{2}} |\boldsymbol{\Gamma}_{\mathbf{X}}|^{\frac{1}{2}}} \int_{-\infty}^{\infty} \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}})^T \boldsymbol{\Gamma}_{\mathbf{X}}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}}) \right] e^{j\boldsymbol{\xi}^T \mathbf{x}} d\mathbf{x}\end{aligned}$$

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



Characteristic Functions

The characteristic function for a Gaussian random vector is

$$\Phi_{\mathbf{X}}(\boldsymbol{\xi}) = \int_{-\infty}^{\infty} f_{\mathbf{X}}(\mathbf{x}) e^{j\boldsymbol{\xi}^T \mathbf{x}} d\mathbf{x}$$

$$= \frac{1}{(2\pi)^{\frac{N}{2}} |\boldsymbol{\Gamma}_{\mathbf{X}}|^{\frac{1}{2}}} \int_{-\infty}^{\infty} \exp \left[-\frac{\mathbf{x}^T \boldsymbol{\Gamma}_{\mathbf{X}}^{-1} \mathbf{x} - 2 \left(\boldsymbol{\mu}_{\mathbf{X}}^T \boldsymbol{\Gamma}_{\mathbf{X}}^{-1} + j\boldsymbol{\xi}^T \right) \mathbf{x} + \boldsymbol{\mu}_{\mathbf{X}}^T \boldsymbol{\Gamma}_{\mathbf{X}}^{-1} \boldsymbol{\mu}_{\mathbf{X}}}{2} \right] d\mathbf{x}$$

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



Characteristic Functions

The characteristic function for a Gaussian random vector is

$$\Phi_{\mathbf{X}}(\boldsymbol{\xi}) = \int_{-\infty}^{\infty} f_{\mathbf{X}}(\mathbf{x}) e^{j\boldsymbol{\xi}^T \mathbf{x}} d\mathbf{x}$$

$$= \frac{1}{(2\pi)^{\frac{N}{2}} |\boldsymbol{\Gamma}_{\mathbf{X}}|^{\frac{1}{2}}} \int_{-\infty}^{\infty} \exp \left[-\frac{\mathbf{x}^T \boldsymbol{\Gamma}_{\mathbf{X}}^{-1} \mathbf{x} - 2 \left(\boldsymbol{\mu}_{\mathbf{X}}^T \boldsymbol{\Gamma}_{\mathbf{X}}^{-1} + j\boldsymbol{\xi}^T \right) \mathbf{x} + \boldsymbol{\mu}_{\mathbf{X}}^T \boldsymbol{\Gamma}_{\mathbf{X}}^{-1} \boldsymbol{\mu}_{\mathbf{X}}}{2} \right] d\mathbf{x}$$

Using the integral identity:

$$\int_{\mathbb{R}^P} \exp \left\{ -\frac{1}{2} [\alpha + 2\mathbf{y}^T \boldsymbol{\beta} + \mathbf{y}^T \boldsymbol{\Gamma} \mathbf{y}] \right\} d\mathbf{y} = \frac{(2\pi)^{\frac{P}{2}}}{|\boldsymbol{\Gamma}|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} [\alpha - \boldsymbol{\beta}^T \boldsymbol{\Gamma}^{-1} \boldsymbol{\beta}] \right\}$$

gives:

$$\Phi_{\mathbf{X}}(\boldsymbol{\xi}) = \exp \left[j\boldsymbol{\xi}^T \boldsymbol{\mu}_{\mathbf{X}} - \frac{1}{2} \boldsymbol{\xi}^T \boldsymbol{\Gamma}_{\mathbf{X}} \boldsymbol{\xi} \right]$$

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Sum of Independent Random Variables

Theorem (Sum of Random Variables and Vectors). If $\mathbf{X}(\zeta)$ and $\mathbf{Y}(\zeta)$ have joint-pdf, $f_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y})$, then $\mathbf{Z}(\zeta) = \mathbf{X}(\zeta) + \mathbf{Y}(\zeta)$ has density function:

$$f_{\mathbf{Z}}(\mathbf{z}) \triangleq f_{\mathbf{X}+\mathbf{Y}}(\mathbf{z}) = \int_{\mathbb{R}} f_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{z} - \mathbf{x}) d\mathbf{x} \quad \diamond$$



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Sum of Independent Random Variables

Theorem (Sum of Random Variables and Vectors). If $\mathbf{X}(\zeta)$ and $\mathbf{Y}(\zeta)$ have joint-pdf, $f_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y})$, then $\mathbf{Z}(\zeta) = \mathbf{X}(\zeta) + \mathbf{Y}(\zeta)$ has density function:

$$f_{\mathbf{Z}}(\mathbf{z}) \triangleq f_{\mathbf{X}+\mathbf{Y}}(\mathbf{z}) = \int_{\mathbb{R}} f_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{z} - \mathbf{x}) d\mathbf{x} \quad \diamond$$

Theorem (Sum of Independent Random Variables and Vectors). If $\mathbf{X}(\zeta)$ and $\mathbf{Y}(\zeta)$ are independent, this result becomes

$$\begin{aligned} f_{\mathbf{Z}}(\mathbf{z}) \triangleq f_{\mathbf{X}+\mathbf{Y}}(\mathbf{z}) &= \int_{\mathbb{R}} f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{Y}}(\mathbf{z} - \mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbb{R}} f_{\mathbf{X}}(\mathbf{z} - \mathbf{y}) f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} = f_{\mathbf{X}}(\mathbf{z}) * f_{\mathbf{Y}}(\mathbf{y}) \diamond \end{aligned}$$



Sum of Independent Random Variables

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- **Sum of Independent Random Variables**
- Central limit theorem

Estimation Theory

MonteCarlo

Theorem (Sum of Random Variables and Vectors). If $\mathbf{X}(\zeta)$ and $\mathbf{Y}(\zeta)$ have joint-pdf, $f_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y})$, then $\mathbf{Z}(\zeta) = \mathbf{X}(\zeta) + \mathbf{Y}(\zeta)$ has density function:

$$f_{\mathbf{Z}}(\mathbf{z}) \triangleq f_{\mathbf{X}+\mathbf{Y}}(\mathbf{z}) = \int_{\mathbb{R}} f_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{z} - \mathbf{x}) d\mathbf{x} \quad \diamond$$

Theorem (Sum of Independent Random Variables and Vectors). If $\mathbf{X}(\zeta)$ and $\mathbf{Y}(\zeta)$ are independent, this result becomes

$$\begin{aligned} f_{\mathbf{Z}}(\mathbf{z}) \triangleq f_{\mathbf{X}+\mathbf{Y}}(\mathbf{z}) &= \int_{\mathbb{R}} f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{Y}}(\mathbf{z} - \mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbb{R}} f_{\mathbf{X}}(\mathbf{z} - \mathbf{y}) f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} = f_{\mathbf{X}}(\mathbf{z}) * f_{\mathbf{Y}}(\mathbf{y}) \diamond \end{aligned}$$

Independent RVs can also be dealt with using **characteristic functions** as introduced in the lecture on scalar random variables.



Sum of Independent Random Variables

If $Z(\zeta) = X(\zeta) + Y(\zeta)$, then its characteristic function is given by:

$$\Phi_Z(\xi) \triangleq \mathbb{E} \left[e^{j\xi Z(\zeta)} \right] = \mathbb{E} \left[e^{j\xi [X(\zeta) + Y(\zeta)]} \right] = \mathbb{E} \left[e^{j\xi X(\zeta)} \right] \mathbb{E} \left[e^{j\xi Y(\zeta)} \right]$$

where the last inequality follows from independence.

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- **Sum of Independent Random Variables**
- Central limit theorem

Estimation Theory

MonteCarlo

Sum of Independent Random Variables

If $Z(\zeta) = X(\zeta) + Y(\zeta)$, then its characteristic function is given by:

$$\Phi_Z(\xi) \triangleq \mathbb{E} \left[e^{j\xi Z(\zeta)} \right] = \mathbb{E} \left[e^{j\xi [X(\zeta) + Y(\zeta)]} \right] = \mathbb{E} \left[e^{j\xi X(\zeta)} \right] \mathbb{E} \left[e^{j\xi Y(\zeta)} \right]$$

where the last inequality follows from independence.

Hence, from the convolution property of the Fourier transform,

$$f_Z(z) = f_X(x) * f_Y(y)$$



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- **Sum of Independent Random Variables**
- Central limit theorem

Estimation Theory

MonteCarlo

Sum of Independent Random Variables

If $Z(\zeta) = X(\zeta) + Y(\zeta)$, then its characteristic function is given by:

$$\Phi_Z(\xi) \triangleq \mathbb{E} \left[e^{j\xi Z(\zeta)} \right] = \mathbb{E} \left[e^{j\xi [X(\zeta) + Y(\zeta)]} \right] = \mathbb{E} \left[e^{j\xi X(\zeta)} \right] \mathbb{E} \left[e^{j\xi Y(\zeta)} \right]$$

where the last inequality follows from independence.

Hence, from the convolution property of the Fourier transform,

$$f_Z(z) = f_X(x) * f_Y(y)$$

This result can be generalised to the summation of M independent RVs:

$$Y(\zeta) = \sum_{k=1}^M c_k X_k(\zeta)$$

where $\{c_k\}_1^M$ is a set of fixed (deterministic) coefficients.



Sum of Independent Random Variables

It follows straightforwardly that:

$$\Phi_Y(\xi) \triangleq \mathbb{E} \left[e^{j\xi Y(\zeta)} \right] = \prod_{k=1}^M \mathbb{E} \left[e^{j\xi c_k X_k(\zeta)} \right] = \prod_{k=1}^M \Phi_{X_k}(c_k \xi)$$

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



Sum of Independent Random Variables

It follows straightforwardly that:

$$\Phi_Y(\xi) \triangleq \mathbb{E} \left[e^{j\xi Y(\zeta)} \right] = \prod_{k=1}^M \mathbb{E} \left[e^{j\xi c_k X_k(\zeta)} \right] = \prod_{k=1}^M \Phi_{X_k}(c_k \xi)$$

Hence, the pdf of $Y(\zeta)$ is given by:

$$f_Y(y) = \frac{1}{|c_1|} f_{X_1} \left(\frac{y}{c_1} \right) * \frac{1}{|c_2|} f_{X_2} \left(\frac{y}{c_2} \right) * \dots * \frac{1}{|c_M|} f_{X_M} \left(\frac{y}{c_M} \right)$$

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



Sum of Independent Random Variables

It follows straightforwardly that:

$$\Phi_Y(\xi) \triangleq \mathbb{E} \left[e^{j\xi Y(\zeta)} \right] = \prod_{k=1}^M \mathbb{E} \left[e^{j\xi c_k X_k(\zeta)} \right] = \prod_{k=1}^M \Phi_{X_k}(c_k \xi)$$

Hence, the pdf of $Y(\zeta)$ is given by:

$$f_Y(y) = \frac{1}{|c_1|} f_{X_1} \left(\frac{y}{c_1} \right) * \frac{1}{|c_2|} f_{X_2} \left(\frac{y}{c_2} \right) * \dots * \frac{1}{|c_M|} f_{X_M} \left(\frac{y}{c_M} \right)$$

Finally, the cumulant generating, or second characteristic, function can be used to determine the n th-order cumulants for $Y(\zeta)$.

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- **Sum of Independent Random Variables**
- Central limit theorem

Estimation Theory

MonteCarlo



Sum of Independent Random Variables

It follows straightforwardly that:

$$\Phi_Y(\xi) \triangleq \mathbb{E} \left[e^{j\xi Y(\zeta)} \right] = \prod_{k=1}^M \mathbb{E} \left[e^{j\xi c_k X_k(\zeta)} \right] = \prod_{k=1}^M \Phi_{X_k}(c_k \xi)$$

Hence, the pdf of $Y(\zeta)$ is given by:

$$f_Y(y) = \frac{1}{|c_1|} f_{X_1} \left(\frac{y}{c_1} \right) * \frac{1}{|c_2|} f_{X_2} \left(\frac{y}{c_2} \right) * \dots * \frac{1}{|c_M|} f_{X_M} \left(\frac{y}{c_M} \right)$$

Finally, the cumulant generating, or second characteristic, function can be used to determine the n th-order cumulants for $Y(\zeta)$.

Recall that

$$\Psi_X(\xi) \triangleq \ln \Phi_X(\xi) = \ln \mathbb{E} \left[e^{j\xi X(\zeta)} \right]$$

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- **Sum of Independent Random Variables**
- Central limit theorem

Estimation Theory

MonteCarlo



Sum of Independent Random Variables

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

$$\Psi_Y(\xi) \triangleq \ln \mathbb{E} \left[e^{j\xi Y(\zeta)} \right] = \sum_{k=1}^M \ln \mathbb{E} \left[e^{j\xi c_k X_k(\zeta)} \right] = \sum_{k=1}^M \Psi_{X_k}(c_k \xi)$$



Sum of Independent Random Variables

$$\Psi_Y(\xi) \triangleq \ln \mathbb{E} \left[e^{j\xi Y(\zeta)} \right] = \sum_{k=1}^M \ln \mathbb{E} \left[e^{j\xi c_k X_k(\zeta)} \right] = \sum_{k=1}^M \Psi_{X_k}(c_k \xi)$$

Therefore, it can readily be shown that the cumulants of $Y(\zeta)$ are given by:

$$\kappa_Y^{(n)} = \sum_{k=1}^M c_k^n \kappa_{X_k}^{(n)}$$

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



Sum of Independent Random Variables

$$\Psi_Y(\xi) \triangleq \ln \mathbb{E} \left[e^{j\xi Y(\zeta)} \right] = \sum_{k=1}^M \ln \mathbb{E} \left[e^{j\xi c_k X_k(\zeta)} \right] = \sum_{k=1}^M \Psi_{X_k}(c_k \xi)$$

Therefore, it can readily be shown that the cumulants of $Y(\zeta)$ are given by:

$$\kappa_Y^{(n)} = \sum_{k=1}^M c_k^n \kappa_{X_k}^{(n)}$$

When these results are extended to the sum of an infinite number of statistically independent random variables, a powerful theorem known as the central limit theorem (CLT) is obtained.

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



Sum of Independent Random Variables

$$\Psi_Y(\xi) \triangleq \ln \mathbb{E} \left[e^{j\xi Y(\zeta)} \right] = \sum_{k=1}^M \ln \mathbb{E} \left[e^{j\xi c_k X_k(\zeta)} \right] = \sum_{k=1}^M \Psi_{X_k}(c_k \xi)$$

Therefore, it can readily be shown that the cumulants of $Y(\zeta)$ are given by:

$$\kappa_Y^{(n)} = \sum_{k=1}^M c_k^n \kappa_{X_k}^{(n)}$$

When these results are extended to the sum of an infinite number of statistically independent random variables, a powerful theorem known as the central limit theorem (CLT) is obtained.

Another interesting concept develops when the sum of independent and identically distributed (i. i. d.) random variables preserve their distribution, which results in so-called **stable distributions**.

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



Central limit theorem

Consider the random variable $Y(\zeta)$ given by:

$$Y_M(\zeta) = \sum_{k=1}^M c_k X_k(\zeta)$$

What is the distribution of $Y_M(\zeta)$ as $M \rightarrow \infty$?

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo



Central limit theorem

Consider the random variable $Y(\zeta)$ given by:

$$Y_M(\zeta) = \sum_{k=1}^M c_k X_k(\zeta)$$

What is the distribution of $Y_M(\zeta)$ as $M \rightarrow \infty$?

Theorem (Central limit theorem). Let $\{X_k(\zeta)\}_{k=1}^M$ be a collection of RVs that are independent and identically distributed for all $k = \{1, \dots, M\}$. Then the distribution of

$$\hat{Y}_M(\zeta) = \frac{Y_M(\zeta) - \mu_{Y_M}}{\sigma_{Y_M}} \quad \text{where} \quad Y_M(\zeta) = \sum_{k=1}^M X_k(\zeta)$$

approaches

$$\lim_{M \rightarrow \infty} f_{\hat{Y}_M}(y) = \mathcal{N}(y | 0, 1)$$



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

- Abstract
- Definition of Random Vectors
- Distribution and Density Functions
- Marginal Density Function
- Independence
- Conditionals and Bayes's
- Statistical Description
- Probability Transformation Rule
- Polar Transformation
- Linear Transformations
- Auxiliary Variables
- Multivariate Gaussian Density Function
- Characteristic Functions
- Sum of Independent Random Variables
- Central limit theorem

Estimation Theory

MonteCarlo

Handout 5

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.

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

● Introduction

- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the maximum-likelihood estimate (MLE)
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

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.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- **Properties of Estimators**
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

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.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- **Properties of Estimators**
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

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.



Properties of Estimators

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- **Properties of Estimators**
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

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:

$$\hat{\theta} = \hat{\theta}[\mathcal{X}] = \hat{\theta}[\{x[n]\}_0^{N-1}]$$



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- **Bias of estimator**
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

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:

$$\hat{\theta} = \hat{\theta}[\mathcal{X}] = \hat{\theta}[\{x[n]\}_0^{N-1}]$$

The function $\hat{\theta}[\mathcal{X}]$ is known as an **estimator** whereas the value taken by the estimator, using a particular set of observations, is called a **point-estimate**.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- **Bias of estimator**
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

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:

$$\hat{\theta} = \hat{\theta}[\mathcal{X}] = \hat{\theta}[\{x[n]\}_0^{N-1}]$$

The function $\hat{\theta}[\mathcal{X}]$ 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.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- **Bias of estimator**
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

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:

$$\hat{\theta} = \hat{\theta}[\mathcal{X}] = \hat{\theta}[\{x[n]\}_0^{N-1}]$$

The function $\hat{\theta}[\mathcal{X}]$ 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.

Since $\hat{\theta}$ is a function of a number of particular realisations of a random outcome (or experiment), then it is itself a RV, and thus has a mean and variance.



Bias of estimator

The **bias** of an estimator $\hat{\theta}$ of a parameter θ is defined as:

$$B(\hat{\theta}) \triangleq \mathbb{E} [\hat{\theta}] - \theta$$

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo



Bias of estimator

The **bias** of an estimator $\hat{\theta}$ of a parameter θ is defined as:

$$B(\hat{\theta}) \triangleq \mathbb{E} [\hat{\theta}] - \theta$$

If θ is large, then a small deviation would give what would appear to be a large bias. Thus, the **normalised bias** is often used instead:

$$\epsilon_b(\hat{\theta}) \triangleq \frac{B(\hat{\theta})}{\theta} = \frac{\mathbb{E} [\hat{\theta}]}{\theta} - 1, \quad \theta \neq 0$$

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo



Bias of estimator

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

The **bias** of an estimator $\hat{\theta}$ of a parameter θ is defined as:

$$B(\hat{\theta}) \triangleq \mathbb{E} [\hat{\theta}] - \theta$$

If θ is large, then a small deviation would give what would appear to be a large bias. Thus, the **normalised bias** is often used instead:

$$\epsilon_b(\hat{\theta}) \triangleq \frac{B(\hat{\theta})}{\theta} = \frac{\mathbb{E} [\hat{\theta}]}{\theta} - 1, \quad \theta \neq 0$$

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

$$\mathbb{E} [\hat{\mu}_x] = \mathbb{E} \left[\frac{1}{N} \sum_{n=0}^{N-1} x[n] \right] = \frac{1}{N} \sum_{n=0}^{N-1} \mathbb{E} [x[n]] = \frac{N\mu_X}{N} = \mu_X.$$



Variance of estimator

The variance of the estimator $\hat{\theta}$ is defined by:

$$\text{var} [\hat{\theta}] = \sigma_{\hat{\theta}}^2 \triangleq \mathbb{E} \left[\left| \hat{\theta} - \mathbb{E} [\hat{\theta}] \right|^2 \right]$$

However, a minimum variance criterion is not always compatible with the minimum bias requirement; reducing the variance may result in an increase in bias.

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

Variance of estimator

The variance of the estimator $\hat{\theta}$ is defined by:

$$\text{var} [\hat{\theta}] = \sigma_{\hat{\theta}}^2 \triangleq \mathbb{E} \left[\left| \hat{\theta} - \mathbb{E} [\hat{\theta}] \right|^2 \right]$$

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 below.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

Variance of estimator

The variance of the estimator $\hat{\theta}$ is defined by:

$$\text{var} [\hat{\theta}] = \sigma_{\hat{\theta}}^2 \triangleq \mathbb{E} \left[\left| \hat{\theta} - \mathbb{E} [\hat{\theta}] \right|^2 \right]$$

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 below.

The normalised standard deviation is defined by:

$$\epsilon_r \triangleq \frac{\sigma_{\hat{\theta}}}{\theta}, \quad \theta \neq 0$$



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- **Cramer-Rao Lower Bound**
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

Mean square error

Minimising estimator variance can increase bias. A compromise criterion is the mean-squared error (MSE) of the estimator, which is given by:

$$\text{MSE}(\hat{\theta}) = \mathbb{E} \left[\left| \hat{\theta} - \theta \right|^2 \right] = \sigma_{\hat{\theta}}^2 + |B(\hat{\theta})|^2$$



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- **Cramer-Rao Lower Bound**
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

Mean square error

Minimising estimator variance can increase bias. A compromise criterion is the mean-squared error (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:

$$\hat{\theta}_{\text{MSE}} = \arg_{\hat{\theta}} \min \text{MSE}(\hat{\theta})$$



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- **Cramer-Rao Lower Bound**
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

Mean square error

Minimising estimator variance can increase bias. A compromise criterion is the mean-squared error (MSE) of the estimator, which is given by:

$$\text{MSE}(\hat{\theta}) = \mathbb{E} \left[|\hat{\theta} - \theta|^2 \right] = \sigma_{\hat{\theta}}^2 + |B(\hat{\theta})|^2$$

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:

$$\hat{\theta}_{\text{MSE}} = \arg_{\hat{\theta}} \min \text{MSE}(\hat{\theta})$$

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.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- **Cramer-Rao Lower Bound**
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

Cramer-Rao Lower Bound

If the MSE can be minimised when the bias is zero, then clearly the variance is also minimised. Such estimators are called minimum variance unbiased estimators (MVUEs).



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- **Cramer-Rao Lower Bound**
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

Cramer-Rao Lower Bound

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).



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- **Cramer-Rao Lower Bound**
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

Cramer-Rao Lower Bound

Theorem (CRLB - scalar parameter). If

$\mathbf{X}(\zeta) = [x[0, \zeta], \dots, x[N - 1, \zeta]]^T$ and $f_{\mathbf{X}}(\mathbf{x} | \theta)$ is the joint density of $\mathbf{X}(\zeta)$ which depends on fixed but unknown parameter θ , then the variance of the estimator $\hat{\theta}$ is bounded by:

$$\text{var} [\hat{\theta}] \geq \frac{1}{\mathbb{E} \left[\left(\frac{\partial \ln f_{\mathbf{X}}(\mathbf{x} | \theta)}{\partial \theta} \right)^2 \right]}$$





Cramer-Rao Lower Bound

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

Theorem (CRLB - scalar parameter). If $\mathbf{X}(\zeta) = [x[0, \zeta], \dots, x[N - 1, \zeta]]^T$ and $f_{\mathbf{X}}(\mathbf{x} | \theta)$ is the joint density of $\mathbf{X}(\zeta)$ which depends on fixed but unknown parameter θ , then the variance of the estimator $\hat{\theta}$ is bounded by:

$$\text{var} [\hat{\theta}] \geq \frac{1}{\mathbb{E} \left[\left(\frac{\partial \ln f_{\mathbf{X}}(\mathbf{x} | \theta)}{\partial \theta} \right)^2 \right]}$$

Alternatively, it may also be expressed as:

$$\text{var} [\hat{\theta}] \geq - \frac{1}{\mathbb{E} \left[\frac{\partial^2 \ln f_{\mathbf{X}}(\mathbf{x} | \theta)}{\partial \theta^2} \right]}$$





Cramer-Rao Lower Bound

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

Theorem (CRLB - scalar parameter). If

$\mathbf{X}(\zeta) = [x[0, \zeta], \dots, x[N - 1, \zeta]]^T$ and $f_{\mathbf{X}}(\mathbf{x} | \theta)$ is the joint density of $\mathbf{X}(\zeta)$ which depends on fixed but unknown parameter θ , then the variance of the estimator $\hat{\theta}$ is bounded by:

$$\text{var} [\hat{\theta}] \geq \frac{1}{\mathbb{E} \left[\left(\frac{\partial \ln f_{\mathbf{X}}(\mathbf{x} | \theta)}{\partial \theta} \right)^2 \right]}$$

Alternatively, it may also be expressed as:

$$\text{var} [\hat{\theta}] \geq - \frac{1}{\mathbb{E} \left[\frac{\partial^2 \ln f_{\mathbf{X}}(\mathbf{x} | \theta)}{\partial \theta^2} \right]}$$



The function $\ln f_{\mathbf{X}}(\mathbf{x} | \theta)$ is called the **log-likelihood** of θ .



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

Cramer-Rao Lower Bound

Theorem (CRLB - scalar parameter). If

$\mathbf{X}(\zeta) = [x[0, \zeta], \dots, x[N - 1, \zeta]]^T$ and $f_{\mathbf{X}}(\mathbf{x} | \theta)$ is the joint density of $\mathbf{X}(\zeta)$ which depends on fixed but unknown parameter θ , then the variance of the estimator $\hat{\theta}$ is bounded by:

$$\text{var} [\hat{\theta}] \geq \frac{1}{\mathbb{E} \left[\left(\frac{\partial \ln f_{\mathbf{X}}(\mathbf{x} | \theta)}{\partial \theta} \right)^2 \right]}$$

Alternatively, it may also be expressed as:

$$\text{var} [\hat{\theta}] \geq - \frac{1}{\mathbb{E} \left[\frac{\partial^2 \ln f_{\mathbf{X}}(\mathbf{x} | \theta)}{\partial \theta^2} \right]}$$

Furthermore, an unbiased estimator may be found that attains the bound for all θ if, and only if, (iff)

$$\frac{\partial \ln f_{\mathbf{X}}(\mathbf{x} | \theta)}{\partial \theta} = I(\theta) (\hat{\theta} - \theta)$$





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.

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

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 $N \rightarrow \infty$, it will become an impulse at θ .



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- **Estimating Multiple Parameters**
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

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 $N \rightarrow \infty$, it will become an impulse at θ .

This is a very important and desirable property, and such an estimator is called a **consistent estimator**.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- **Maximum Likelihood Estimation**
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

Estimating Multiple Parameters

Multiple parameters occur in, for example, estimating the statistical properties of a random time-series, estimating the parameters of a curve fitted to a set of data, estimating any model described by a set of parameters.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- **Maximum Likelihood Estimation**
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

Estimating Multiple Parameters

Multiple parameters occur in, for example, estimating the statistical properties of a random time-series, estimating the parameters of a curve fitted to a set of data, estimating any model described by a set of parameters.

A vector of parameters, θ , of a random event $X(\zeta)$ can be estimated from a set of observations, $\mathcal{X} = \{x[n]\}_0^{N-1}$, using some function:

$$\hat{\theta} = \hat{\theta}[\mathcal{X}] = \hat{\theta}[\{x[n]\}_0^{N-1}]$$



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- **Maximum Likelihood Estimation**
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

Estimating Multiple Parameters

Multiple parameters occur in, for example, estimating the statistical properties of a random time-series, estimating the parameters of a curve fitted to a set of data, estimating any model described by a set of parameters.

A vector of parameters, θ , of a random event $X(\zeta)$ can be estimated from a set of observations, $\mathcal{X} = \{x[n]\}_0^{N-1}$, using some function:

$$\hat{\theta} = \hat{\theta}[\mathcal{X}] = \hat{\theta}[\{x[n]\}_0^{N-1}]$$

The definitions of **unbiasedness**, **consistency**, and the **CRLB** are all straightforward extensions of the definitions and results for scalar parameter estimates.



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 $f_{\mathbf{X}}(\mathbf{x} | \boldsymbol{\theta})$.

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- **Maximum Likelihood Estimation**
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

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 θ , is $f_{\mathbf{X}}(\mathbf{x} | \theta)$.

This same quantity, viewed as a function of the parameter θ when a particular set of observations, $\hat{\mathbf{x}}$ is given, is known as the **likelihood function**.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

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 θ , is $f_{\mathbf{X}}(\mathbf{x} | \theta)$.

This same quantity, viewed as a function of the parameter θ 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 θ , denoted by $\hat{\theta}_{ml}$, is defined as that value of θ that maximises $f_{\mathbf{X}}(\hat{\mathbf{x}} | \theta)$.



Maximum Likelihood Estimation

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- **Properties of the MLE**
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

The joint density of the RVs $\mathbf{X}(\zeta) = \{x[n, \zeta]\}_0^{N-1}$, which depends on fixed but unknown parameter θ , is $f_{\mathbf{X}}(\mathbf{x} | \theta)$.

This same quantity, viewed as a function of the parameter θ 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 θ , denoted by $\hat{\theta}_{ml}$, is defined as that value of θ that maximises $f_{\mathbf{X}}(\hat{\mathbf{x}} | \theta)$.

The MLE for θ is defined by:

$$\hat{\theta}_{ml}(\mathbf{x}) = \arg_{\theta} \max f_{\mathbf{X}}(\mathbf{x} | \theta)$$

Note that since $\hat{\theta}_{ml}(\mathbf{x})$ depends on the random observation vector \mathbf{x} , and so is *itself a RV*.



Properties of the MLE

1. The MLE satisfies

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

$$\nabla_{\theta} f_{\mathbf{X}}(\mathbf{x} | \theta) \Big|_{\theta = \hat{\theta}_{ml}} = \mathbf{0}_{P \times 1}$$
$$\nabla_{\theta} \ln f_{\mathbf{X}}(\mathbf{x} | \theta) \Big|_{\theta = \hat{\theta}_{ml}} = \mathbf{0}_{P \times 1}$$



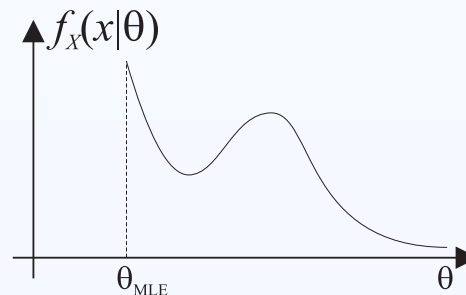
Properties of the MLE

1. The MLE satisfies

$$\nabla_{\theta} f_{\mathbf{X}}(\mathbf{x} | \theta) |_{\theta = \hat{\theta}_{ml}} = \mathbf{0}_{P \times 1}$$

$$\nabla_{\theta} \ln f_{\mathbf{X}}(\mathbf{x} | \theta) |_{\theta = \hat{\theta}_{ml}} = \mathbf{0}_{P \times 1}$$

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

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo



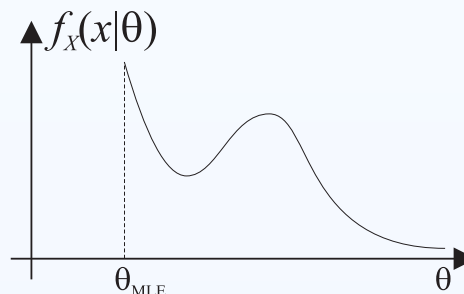
Properties of the MLE

1. The MLE satisfies

$$\nabla_{\theta} f_{\mathbf{X}}(\mathbf{x} | \theta) |_{\theta = \hat{\theta}_{ml}} = \mathbf{0}_{P \times 1}$$

$$\nabla_{\theta} \ln f_{\mathbf{X}}(\mathbf{x} | \theta) |_{\theta = \hat{\theta}_{ml}} = \mathbf{0}_{P \times 1}$$

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. If the pdf, $f_{\mathbf{X}}(\mathbf{x} | \theta)$, satisfies certain *regularity* conditions, then the MLE is asymptotically distributed according to a Gaussian distribution:

$$\hat{\theta}_{ml} \sim \mathcal{N}(\theta, \mathbf{J}^{-1}(\theta))$$

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- **MLE for Transformed Parameter**
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

DC Level in white Gaussian noise

Example ([Therrien:1991, Example 6.1, Page 282]). A constant but unknown signal is observed in additive white Gaussian noise (WGN). That is,

$$x[n] = A + w[n] \quad \text{where} \quad w[n] \sim \mathcal{N}(0, \sigma_w^2) \quad \boxtimes$$

for $n \in \mathcal{N} = \{0, \dots, N - 1\}$. Calculate the MLE of the unknown signal A .



DC Level in white Gaussian noise

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

Example ([Therrien:1991, Example 6.1, Page 282]). A constant but unknown signal is observed in additive WGN. That is,

$$x[n] = A + w[n] \quad \text{where} \quad w[n] \sim \mathcal{N}(0, \sigma_w^2)$$

for $n \in \mathcal{N} = \{0, \dots, N - 1\}$. Calculate the MLE of the unknown signal A .

SOLUTION. Since this is a memoryless system, and $w(n)$ are i. i. d., then so is $x[n]$, and

the log-likelihood is given by:

$$\ln f_{\mathbf{x}}(\mathbf{x} | A) = -\frac{N}{2} \ln(2\pi\sigma_w^2) - \frac{\sum_{n \in \mathcal{N}} (x[n] - A)^2}{2\sigma_w^2}$$

Differentiating this expression w. r. t. A

and setting to zero :

$$\hat{A} = \frac{1}{N} \sum x[n]$$



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- **Least Squares**
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

MLE for Transformed Parameter

Theorem (Invariance Property of the MLE). The MLE of the parameter $\alpha = g(\theta)$, where g is an r -dimensional function of the $P \times 1$ parameter θ , and the pdf, $f_{\mathbf{X}}(\mathbf{x} | \theta)$ is parameterised by θ , is given by

$$\hat{\alpha}_{ml} = g(\hat{\theta}_{ml})$$



where $\hat{\theta}_{ml}$ is the MLE of θ .



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

MLE for Transformed Parameter

Theorem (Invariance Property of the MLE). The MLE of the parameter $\alpha = g(\theta)$, where g is an r -dimensional function of the $P \times 1$ parameter θ , and the pdf, $f_{\mathbf{X}}(\mathbf{x} | \theta)$ is parameterised by θ , is given by

$$\hat{\alpha}_{ml} = g(\hat{\theta}_{ml})$$

where $\hat{\theta}_{ml}$ is the MLE of θ .

The MLE of θ , $\hat{\theta}_{ml}$, is obtained by maximising $f_{\mathbf{X}}(\mathbf{x} | \theta)$. If the function g is not an invertible function, then $\hat{\alpha}$ maximises the modified likelihood function $\bar{p}_T(\mathbf{x} | \alpha)$ defined as:

$$\bar{p}_T(\mathbf{x} | \alpha) = \max_{\theta: \alpha=g(\theta)} f_{\mathbf{X}}(\mathbf{x} | \theta)$$





Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- **Least Squares**
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

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.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- **The Least Squares Approach**
- DC Level
- Linear Least Squares

MonteCarlo

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.

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 of interest: the **principle of least squares**.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- **The Least Squares Approach**
- DC Level
- Linear Least Squares

MonteCarlo

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.

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 of interest: 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.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- **The Least Squares Approach**
- DC Level
- Linear Least Squares

MonteCarlo

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.

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 of interest: 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 least-squares estimate (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.

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- **The Least Squares Approach**
- DC Level
- Linear Least Squares

MonteCarlo



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

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 hidden or unobserved signal is generated by some model which, in turn, depends on some unknown parameter θ .



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

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 hidden or unobserved signal is generated by some model which, in turn, depends on some unknown parameter θ .

The LSE of θ chooses the value that makes $s[n]$ closest to the observed data $x[n]$, and this *closeness* is measured by the LS error criterion:

$$J(\theta) = \sum_{n=0}^{N-1} (x[n] - s[n])^2$$



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

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 hidden or unobserved signal is generated by some model which, in turn, depends on some unknown parameter θ .

The LSE of θ chooses the value that makes $s[n]$ closest to the observed data $x[n]$, and this *closeness* is measured by the LS error criterion:

$$J(\theta) = \sum_{n=0}^{N-1} (x[n] - s[n])^2$$

The LSE is given by:

$$\hat{\theta}_{LSE} = \arg_{\theta} \min J(\theta)$$



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- **Linear Least Squares**

MonteCarlo

DC Level

Example ([Kay:1993, Example 6.1, Page 221]). 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 .



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

DC Level

Example ([Kay:1993, Example 6.1, Page 221]). 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

$$\hat{A}_{LSE} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] \quad \square$$

which is the sample mean estimator.



Linear Least Squares

Thus, the unknown random-vector s is linear in the unknown parameter vector $\theta = [\theta_1, \dots, \theta_P]$,

$$s = \mathbf{H}\theta$$

The LSE is found by minimising:

$$J(\theta) = \sum_{n=0}^{N-1} |x[n] - s[n]|^2 = (\mathbf{x} - \mathbf{H}\theta)^T (\mathbf{x} - \mathbf{H}\theta)$$

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

Linear Least Squares

Thus, the unknown random-vector \mathbf{s} is linear in the unknown parameter vector $\boldsymbol{\theta} = [\theta_1, \dots, \theta_P]$,

$$\mathbf{s} = \mathbf{H}\boldsymbol{\theta}$$

The LSE is found by minimising:

$$J(\boldsymbol{\theta}) = \sum_{n=0}^{N-1} |x[n] - s[n]|^2 = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})$$

Setting the gradient of $J(\boldsymbol{\theta})$ to zero yields the LSE:

$$\hat{\boldsymbol{\theta}}_{LSE} = \left(\mathbf{H}^T \mathbf{H}\right)^{-1} \mathbf{H}^T \mathbf{x}$$



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

- Introduction
- Properties of Estimators
- Bias of estimator
- Variance of estimator
- Mean square error
- Cramer-Rao Lower Bound
- Consistency of an Estimator
- Estimating Multiple Parameters
- Maximum Likelihood Estimation
- Properties of the MLE
- DC Level in white Gaussian noise
- MLE for Transformed Parameter
- Least Squares
- The Least Squares Approach
- DC Level
- Linear Least Squares

MonteCarlo

Linear Least Squares

Thus, the unknown random-vector \mathbf{s} is linear in the unknown parameter vector $\boldsymbol{\theta} = [\theta_1, \dots, \theta_P]$,

$$\mathbf{s} = \mathbf{H}\boldsymbol{\theta}$$

The LSE is found by minimising:

$$J(\boldsymbol{\theta}) = \sum_{n=0}^{N-1} |x[n] - s[n]|^2 = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})$$

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**.

Handout 6

MonteCarlo



Introduction

Many signal processing problems can be reduced to either an *optimisation* problem or an *integration* problem:

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

● Introduction

- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling

Introduction

Many signal processing problems can be reduced to either an *optimisation* problem or an *integration* problem:

Optimisation: involves finding the solution to

$$\hat{\theta} = \arg \max_{\theta \in \Theta} h(\theta)$$

where $h(\cdot)$ is a scalar function of a multi-dimensional vector of parameters, θ .

Typically, $h(\cdot)$ might represent some **cost function**, and it is implicitly assumed that the optimisation cannot be calculated explicitly.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- **Deterministic Numerical Methods**
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling

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*.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- **Deterministic Numerical Methods**
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling

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.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- **Deterministic Numerical Methods**
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling

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:

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \int_{\Omega} f(\theta, \omega) d\omega$$



Deterministic Numerical Methods

Aims and Objectives

Probability Theory

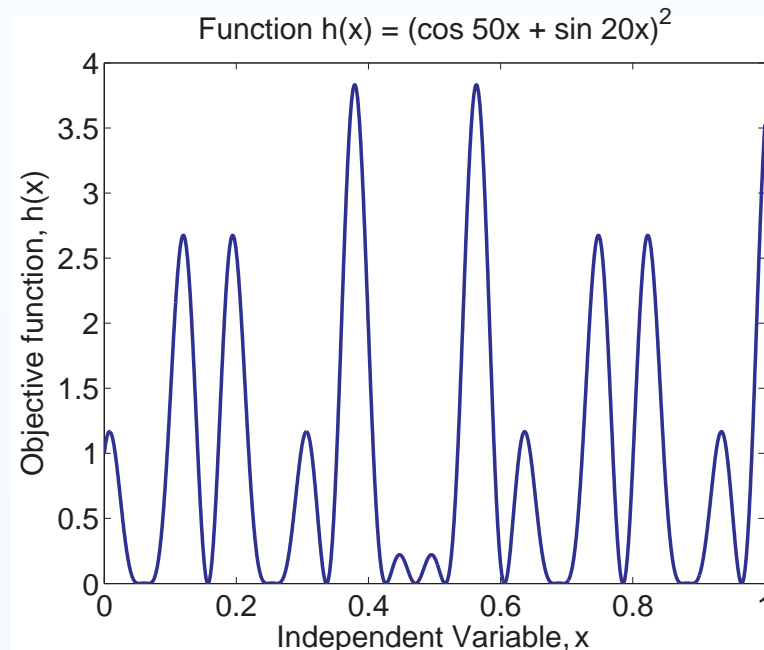
Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- **Deterministic Numerical Methods**
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling



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.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- **Deterministic Optimisation**
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling

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.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- **Deterministic Optimisation**
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling

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.

Unfortunately, these methods are not easily extended to multi-dimensions.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- **Deterministic Optimisation**
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling

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-uniformly 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.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- **Deterministic Integration**
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling

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.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- **Deterministic Integration**
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling

Deterministic Optimisation

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:



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- **Monte Carlo Numerical Methods**
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling

Deterministic Optimisation

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:

$$\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n - (\nabla \nabla^T h(\boldsymbol{\theta}_n))^{-1} \nabla h(\boldsymbol{\theta}_n)$$

Numerous variants of Newton-Raphson-type techniques exist, and include the **steepest descent method**, or the **Levenberg-Marquardt method**.



Deterministic Integration

The integral

$$\mathcal{I} = \int_a^b f(\theta) d\theta,$$

where θ is a scalar, and $b > a$, can be solved with the trapezoidal rule using

$$\hat{I} = \frac{1}{2} \sum_{k=0}^{N-1} (\theta_{k+1} - \theta_k) (f(\theta_k) + f(\theta_{k+1}))$$

where the θ_k 's constitute an ordered partition of $[a, b]$.

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- **Monte Carlo Integration**
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling



Deterministic Integration

The integral

$$\mathcal{I} = \int_a^b f(\theta) d\theta,$$

where θ is a scalar, and $b > a$, can be solved with the trapezoidal rule using

$$\hat{I} = \frac{1}{2} \sum_{k=0}^{N-1} (\theta_{k+1} - \theta_k) (f(\theta_k) + f(\theta_{k+1}))$$

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 $\delta = \theta_{k+1} - \theta_k$.

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- **Monte Carlo Integration**
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling



Monte Carlo Numerical Methods

Monte Carlo methods are stochastic techniques, in which random numbers are generated and use to examine some problem.

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling



Monte Carlo Integration

Consider the integral,

$$\mathcal{I} = \int_{\Theta} f(\boldsymbol{\theta}) d\boldsymbol{\theta}.$$

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- **Monte Carlo Integration**
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling



Monte Carlo Integration

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling

Consider the integral,

$$\mathcal{I} = \int_{\Theta} f(\boldsymbol{\theta}) d\boldsymbol{\theta}.$$

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

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling

Consider the integral,

$$\mathcal{I} = \int_{\Theta} f(\boldsymbol{\theta}) d\boldsymbol{\theta}.$$

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:

$$\mathcal{I} = \mathbb{E}_{\pi} \left[\frac{f(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta})} \right]$$



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling

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

$$\hat{\mathcal{I}} = \frac{1}{N} \sum_{k=0}^{N-1} \frac{f(\boldsymbol{\theta}^{(k)})}{\pi(\boldsymbol{\theta}^{(k)})} \approx \mathbb{E}_{\pi} \left[\frac{f(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta})} \right]$$



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling

Stochastic Optimisation

There are two distinct approaches to the Monte Carlo optimisation of the objective function $h(\theta)$:

$$\hat{\theta} = \arg \max_{\theta \in \Theta} h(\theta)$$

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.

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- **Generating Random Variables**
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- **Uniform Variates**
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling

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:

$$\hat{\boldsymbol{\theta}} \approx \arg \max_{\{\boldsymbol{\theta}^{(k)}\}} h(\boldsymbol{\theta}^{(k)})$$

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 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, $\{\theta^{(k)}\}$, from a pdf $\pi(\theta)$, and taking the estimate:

$$\hat{\theta} \approx \arg \max_{\{\theta^{(k)}\}} h(\theta^{(k)})$$

Typically, when no specific features regarding the function $h(\theta)$, are taken into account, $\pi(\theta)$ will take on a uniform distribution over Θ .

Stochastic Approximation 🎯 The Monte Carlo EM algorithm

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- **Uniform Variates**
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling



Generating Random Variables

This section discusses a variety of techniques for generating random variables from a different distributions.

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
 - **Uniform Variates**
 - Transformation Methods
 - Inverse Transform Method
 - Acceptance-Rejection Sampling
 - Envelope and Squeeze Methods
 - Importance Sampling
 - Other Methods
 - Markov chain Monte Carlo Methods
 - The Metropolis-Hastings algorithm
 - Gibbs Sampling



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling

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 $(0, 1]$.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling

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:

$$x_{n+1} = (a x_n + b) \mod m$$



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling

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.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling

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

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

PROOF. The proof is given in the handout on scalar random variables.



Inverse Transform Method

Aims and Objectives

Probability Theory

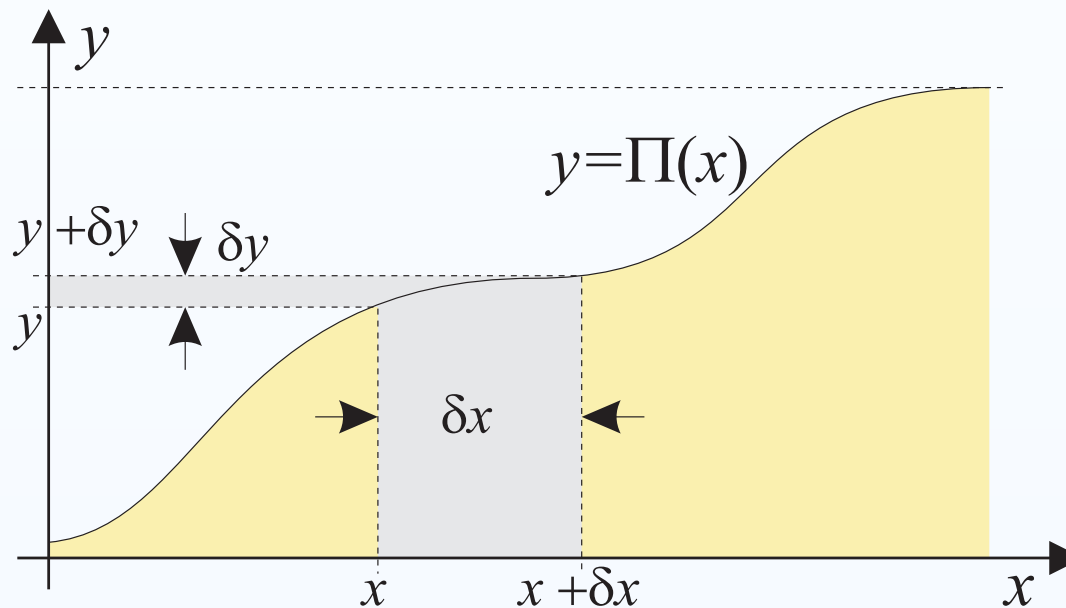
Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- **Inverse Transform Method**
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling



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

Aims and Objectives

Probability Theory

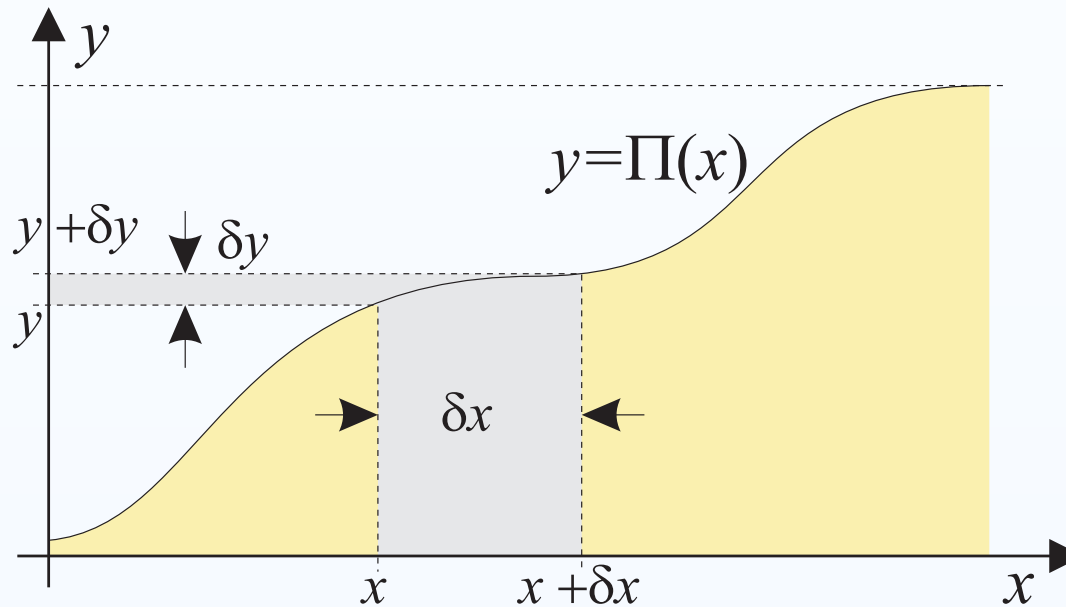
Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling



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

$$f_X(x) = \pi(x), \quad \text{where } x = \Pi^{-1}(y)$$



Inverse Transform Method

In otherwords, if

$$U(\zeta) \sim \mathcal{U}_{[0,1]}, X(\zeta) = \Pi^{-1}U(\zeta) \sim \pi(x)$$

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling



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 $X(\zeta) \sim \mathcal{Exp}(1)$, such that $\pi(x) = e^{-x}$ and $\Pi(x) = 1 - e^{-x}$, then solving for x in terms of $u = 1 - e^{-x}$ gives $x = -\log(1 - u)$.

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling



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 $X(\zeta) \sim \mathcal{Exp}(1)$, 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).

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling

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

Aims and Objectives

Probability Theory

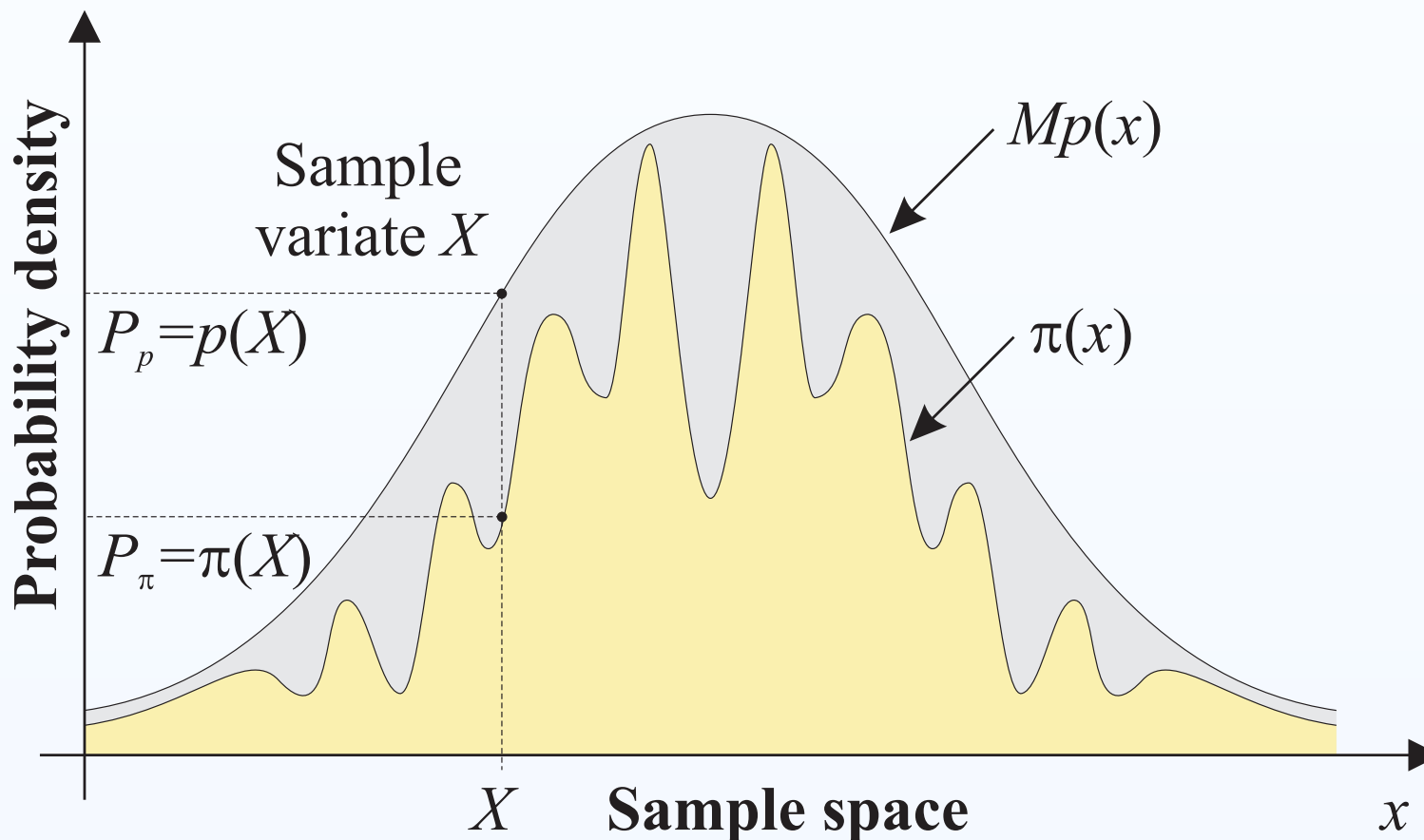
Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling



On average, you would expect to have too many variates that take on the value X by a factor of

$$u(X) = \frac{P_p}{P_\pi} = \frac{p(X)}{\pi(X)}$$



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs 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*.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs 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 $U \leq P_a = \frac{\pi(X)}{Mp(x)}$;
3. Otherwise, reject and return to first step.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling

Envelope and Squeeze Methods

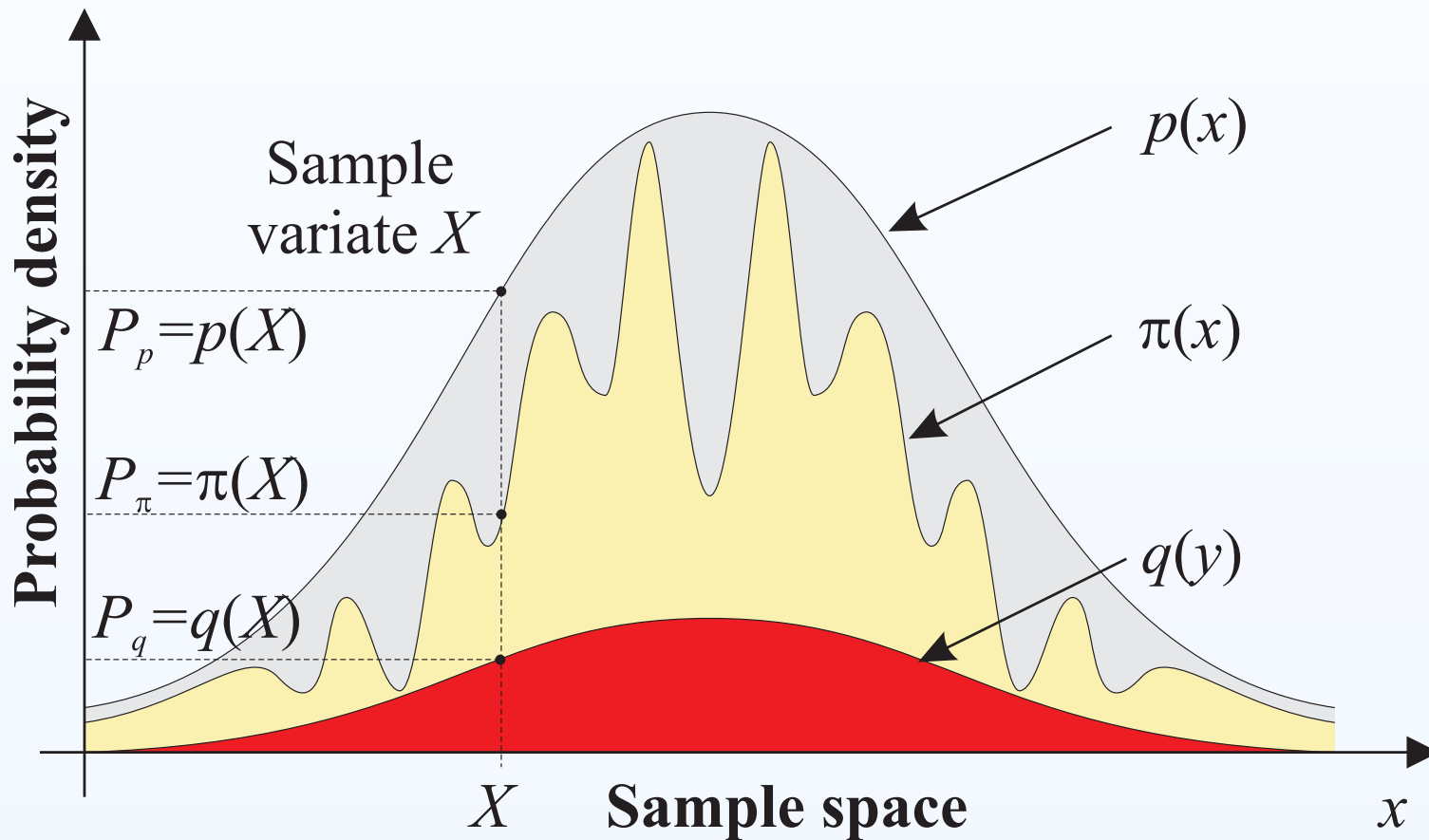
A problem with many sampling methods, which can make the density $\pi(x)$ difficult to simulate, is that the function may require substantial computing time at each evaluation.

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 $U \leq \frac{\pi(X)}{Mp(x)}$.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling

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 $U \leq \frac{q(X)}{Mp(x)}$;
3. Otherwise, accept X if $U \leq \frac{\pi(X)}{Mp(x)}$;
4. Otherwise, reject and return to first step.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling

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 $U \leq \frac{q(X)}{Mp(x)}$;
3. Otherwise, accept X if $U \leq \frac{\pi(X)}{Mp(x)}$;
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 .

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling

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 it is desired to evaluate the function:

$$\mathcal{I} = \int_{\Theta} f(\boldsymbol{\theta}) d\boldsymbol{\theta}.$$



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling

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 it is desired to evaluate the function:

$$\mathcal{I} = \int_{\Theta} f(\boldsymbol{\theta}) d\boldsymbol{\theta}.$$

Approximate by empirical average:

$$\hat{\mathcal{I}} = \frac{1}{N} \sum_{k=0}^{N-1} \mathbb{I}_{\Theta} \left(\boldsymbol{\theta}^{(k)} \right), \quad \text{where } \boldsymbol{\theta}^{(k)} \sim f(\boldsymbol{\theta})$$

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 $\pi(\boldsymbol{\theta}) > 0, \forall \boldsymbol{\theta} \in \Theta$:

$$\mathcal{I} = \int_{\Theta} \frac{f(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta})} \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} = \mathbb{E}_{\pi} \left[\frac{f(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta})} \right],$$

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling



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**;

$$\hat{\mathcal{I}} = \frac{1}{N} \sum_{k=0}^{N-1} \frac{f(\boldsymbol{\theta}^{(k)})}{\pi(\boldsymbol{\theta}^{(k)})}$$

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling

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).



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm

● Gibbs Sampling

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.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm

● Gibbs Sampling

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**:
 $Y \sim g(y | X^{(k)})$.

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

Aims and Objectives

Probability Theory

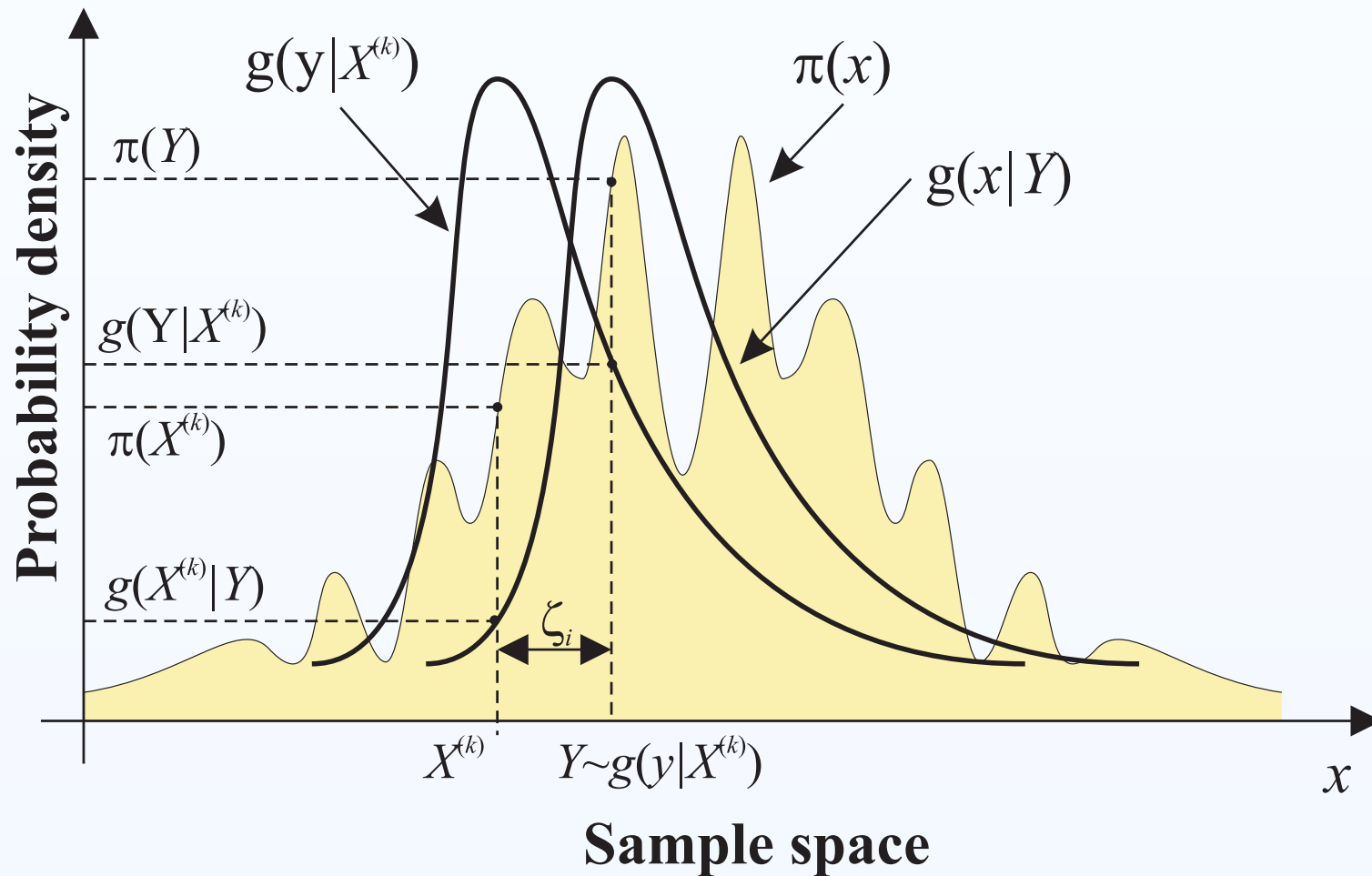
Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm



● Gibbs Sampling



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm

● Gibbs Sampling

Gibbs Sampling

Gibbs sampling is a Monte Carlo method that facilitates sampling from a multivariate density function, $\pi(\theta_0, \theta_1, \dots, \theta_M)$ by drawing successive samples from marginal densities of smaller dimensions.



Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm

● Gibbs Sampling

Gibbs Sampling

Gibbs sampling is a Monte Carlo method that facilitates sampling from a multivariate density function, $\pi(\theta_0, \theta_1, \dots, \theta_M)$ by drawing successive samples from marginal densities of smaller dimensions.

Using the probability chain rule,

$$\pi(\{\theta_m\}_{m=1}^M) = \pi(\theta_\ell | \{\theta_m\}_{m=1, m \neq \ell}^M) \pi(\{\theta_m\}_{m=1, m \neq \ell}^M)$$

The Gibbs sampler works by drawing random variates from the marginal densities $\pi(\theta_\ell | \{\theta_m\}_{m=1, m \neq \ell}^M)$ in a cyclic iterative pattern.



Gibbs Sampling

First iteration:

$$\theta_1^{(1)} \sim \pi \left(\theta_1 \mid \theta_2^{(0)}, \theta_3^{(0)}, \theta_4^{(0)}, \dots, \theta_M^{(0)} \right)$$

$$\theta_2^{(1)} \sim \pi \left(\theta_2 \mid \theta_1^{(1)}, \theta_3^{(0)}, \theta_4^{(0)}, \dots, \theta_M^{(0)} \right)$$

$$\theta_3^{(1)} \sim \pi \left(\theta_3 \mid \theta_1^{(1)}, \theta_2^{(1)}, \theta_4^{(0)}, \dots, \theta_M^{(0)} \right)$$

⋮ ⋮

$$\theta_M^{(1)} \sim \pi \left(\theta_M \mid \theta_1^{(1)}, \theta_2^{(1)}, \theta_4^{(1)}, \dots, \theta_{M-1}^{(1)} \right)$$

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling



Gibbs Sampling

Second iteration:

$$\theta_1^{(2)} \sim \pi \left(\theta_1 \mid \theta_2^{(1)}, \theta_3^{(1)}, \theta_4^{(1)}, \dots, \theta_M^{(1)} \right)$$

$$\theta_2^{(2)} \sim \pi \left(\theta_2 \mid \theta_1^{(2)}, \theta_3^{(1)}, \theta_4^{(1)}, \dots, \theta_M^{(1)} \right)$$

$$\theta_3^{(2)} \sim \pi \left(\theta_3 \mid \theta_1^{(2)}, \theta_2^{(2)}, \theta_4^{(1)}, \dots, \theta_M^{(1)} \right)$$

⋮ ⋮

$$\theta_M^{(2)} \sim \pi \left(\theta_M \mid \theta_1^{(2)}, \theta_2^{(2)}, \theta_4^{(2)}, \dots, \theta_{M-1}^{(2)} \right)$$

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling



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 \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 $\pi(\theta_0, \theta_1, \dots, \theta_M)$.

Aims and Objectives

Probability Theory

Scalar Random Variables

Multiple Random Variables

Estimation Theory

MonteCarlo

- Introduction
- Deterministic Numerical Methods
- Deterministic Optimisation
- Deterministic Integration
- Monte Carlo Numerical Methods
- Monte Carlo Integration
- Stochastic Optimisation
- Generating Random Variables
- Uniform Variates
- Transformation Methods
- Inverse Transform Method
- Acceptance-Rejection Sampling
- Envelope and Squeeze Methods
- Importance Sampling
- Other Methods
- Markov chain Monte Carlo Methods
- The Metropolis-Hastings algorithm
- Gibbs Sampling