Probabilistic Program Analysis

Computation and Probability

Alessandra Di Pierro
University of Verona, Italy
alessandra.dipierro@univr.it

Herbert Wiklicky
Imperial College London, UK
herbert@doc.ic.ac.uk

Practicalities

Two lecturers for this introductory course:

Herbert Wiklicky
h.wiklicky@imperial.ac.uk
Alessandra Di Pierro
alessandra.dipierro@univr.it

Approximate schedule:

- Motivation: Computation and Probability
- Syntax and Semantics of a Probabilistic Language
- Probabilistic Abstract Interpretation
- Probabilistic Data-Flow Analysis
- Logic of PAI, Precision, Applications, etc.

Practicalities

Two lecturers for this introductory course:

Herbert Wiklicky
h.wiklicky@imperial.ac.uk
Alessandra Di Pierro
alessandra.dipierro@univr.it

Approximate schedule:

- Motivation: Computation and Probability
- Syntax and Semantics of a Probabilistic Language
- Probabilistic Abstract Interpretation
- Probabilistic Data-Flow Analysis
- Logic of PAI, Precision, Applications, etc.

Practicalities

Two lecturers for this introductory course:

Herbert Wiklicky

h.wiklicky@imperial.ac.uk

Alessandra Di Pierro

alessandra.dipierro@univr.it

Approximate schedule:

- Motivation: Computation and Probability
- Syntax and Semantics of a Probabilistic Language
- Probabilistic Abstract Interpretation
- Probabilistic Data-Flow Analysis
- Logic of PAI, Precision, Applications, etc.

Practicalities

Two lecturers for this introductory course:

Herbert Wiklicky
h.wiklicky@imperial.ac.uk
Alessandra Di Pierro
alessandra.dipierro@univr.it

Approximate schedule:

- Motivation: Computation and Probability
- Syntax and Semantics of a Probabilistic Language
- Probabilistic Abstract Interpretation
- Probabilistic Data-Flow Analysis
- Logic of PAI, Precision, Applications, etc.

Practicalities

Two lecturers for this introductory course:

Herbert Wiklicky
h.wiklicky@imperial.ac.uk
Alessandra Di Pierro
alessandra.dipierro@univr.it

Approximate schedule:

- Motivation: Computation and Probability
- Syntax and Semantics of a Probabilistic Language
- Probabilistic Abstract Interpretation
- Probabilistic Data-Flow Analysis
- Logic of PAI, Precision, Applications, etc.

Practicalities

Two lecturers for this introductory course:

Herbert Wiklicky
h.wiklicky@imperial.ac.uk
Alessandra Di Pierro
alessandra.dipierro@univr.it

Approximate schedule:

- Motivation: Computation and Probability
- Syntax and Semantics of a Probabilistic Language
- Probabilistic Abstract Interpretation
- Probabilistic Data-Flow Analysis
- Logic of PAI, Precision, Applications, etc.

Practicalities

Two lecturers for this introductory course:

Herbert Wiklicky

h.wiklicky@imperial.ac.uk

Alessandra Di Pierro

alessandra.dipierro@univr.it

Approximate schedule:

- Motivation: Computation and Probability
- Syntax and Semantics of a Probabilistic Language
- Probabilistic Abstract Interpretation
- Probabilistic Data-Flow Analysis
- Logic of PAI, Precision, Applications, etc.

Practicalities

Two lecturers for this introductory course:

Herbert Wiklicky
h.wiklicky@imperial.ac.uk
Alessandra Di Pierro
alessandra.dipierro@univr.it

Approximate schedule:

- Motivation: Computation and Probability
- Syntax and Semantics of a Probabilistic Language
- Probabilistic Abstract Interpretation
- Probabilistic Data-Flow Analysis
- Logic of PAI, Precision, Applications, etc.

Practicalities

Two lecturers for this introductory course:

Herbert Wiklicky

h.wiklicky@imperial.ac.uk

Alessandra Di Pierro

alessandra.dipierro@univr.it

Approximate schedule:

- Motivation: Computation and Probability
- Syntax and Semantics of a Probabilistic Language
- Probabilistic Abstract Interpretation
- Probabilistic Data-Flow Analysis
- Logic of PAI, Precision, Applications, etc.

Probability and Computation

Commonly, computations are understood to follow a well defined (deterministic) set of rules as to obtain a certain result.

There are randomised algorithms which involve an element of chance or randomness.

Las Vegas Algorithms are randomised algorithms that always give correct results (with non-deterministic running time), e.g. QuickSort (with random pivoting).

Monte Carlo Algorithms produce (with deterministic running time) an output which may be incorrect with a certain probability, e.g. Buffon's Needle.

Probability and Computation

Commonly, computations are understood to follow a well defined (deterministic) set of rules as to obtain a certain result.

There are randomised algorithms which involve an element of chance or randomness.

Las Vegas Algorithms are randomised algorithms that always give correct results (with non-deterministic running time), e.g. QuickSort (with random pivoting).

Monte Carlo Algorithms produce (with deterministic running time) an output which may be incorrect with a certain probability, e.g. Buffon's Needle.

Probability and Computation

Commonly, computations are understood to follow a well defined (deterministic) set of rules as to obtain a certain result.

There are randomised algorithms which involve an element of chance or randomness.

Las Vegas Algorithms are randomised algorithms that always give correct results (with non-deterministic running time), e.g. QuickSort (with random pivoting).

Monte Carlo Algorithms produce (with deterministic running time) an output which may be incorrect with a certain probability, e.g. Buffon's Needle.

Probability and Computation

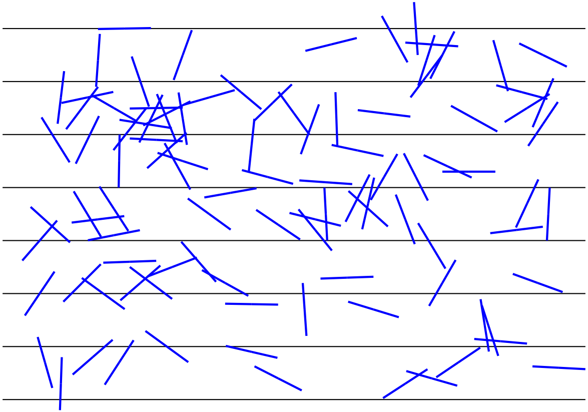
Commonly, computations are understood to follow a well defined (deterministic) set of rules as to obtain a certain result.

There are randomised algorithms which involve an element of chance or randomness.

Las Vegas Algorithms
: are randomised algorithms that always give correct results (with non-deterministic running time), e.g. QuickSort (with random pivoting).

Monte Carlo Algorithms
: produce (with deterministic running time) an output which may be incorrect with a certain probability, e.g. Buffon's Needle.

(Georges-Louis Leclerc, Comte de) Buffon's Needle



$$\Pr(\text{cross}) = \frac{2}{\pi} \text{ or } \pi = \frac{2}{\Pr(\text{cross})}$$

Information and Security

Side-Channel Attacks (Kocher, 1996)
The problem appears in attacks against public encryption algorithms like RSA. In (optimised) versions of de/encoding (using modular exponentation) properties of the secrete key determine the execution time.

How much information about the secret key is revealed?

Differential Privacy (Dwork, 2006)
In large (statistical) databases an attacker can try to reveal information about individuals (de-anonymise), e.g. there are only three under-25 with hair loss registered, and there are two people getting hair-loss treatment in Bolzano. Andrea is on the database, 21 and lives in Bozen.

How much information about individuals is revealed?

Information and Security

Side-Channel Attacks (Kocher, 1996)
The problem appears in attacks against public encryption algorithms like RSA. In (optimised) versions of de/encoding (using modular exponentation) properties of the secrete key determine the execution time.

How much information about the secret key is revealed?

Differential Privacy (Dwork, 2006)
In large (statistical) databases an attacker can try to reveal information about individuals (de-anonymise), e.g. there are only three under-25 with hair loss registered, and there are two people getting hair-loss treatment in Bolzano. Andrea is on the database, 21 and lives in Bozen.

How much information about individuals is revealed?

Information and Security

Side-Channel Attacks (Kocher, 1996)
The problem appears in attacks against public encryption algorithms like RSA. In (optimised) versions of de/encoding (using modular exponentation) properties of the secrete key determine the execution time.

How much information about the secret key is revealed?

Differential Privacy (Dwork, 2006)
In large (statistical) databases an attacker can try to reveal information about individuals (de-anonymise), e.g. there are only three under-25 with hair loss registered, and there are two people getting hair-loss treatment in Bolzano. Andrea is on the database, 21 and lives in Bozen.

How much information about individuals is revealed?

Information and Security

Side-Channel Attacks (Kocher, 1996)
The problem appears in attacks against public encryption algorithms like RSA. In (optimised) versions of de/encoding (using modular exponentation) properties of the secrete key determine the execution time.

How much information about the secret key is revealed?

Differential Privacy (Dwork, 2006)
In large (statistical) databases an attacker can try to reveal information about individuals (de-anonymise), e.g. there are only three under-25 with hair loss registered, and there are two people getting hair-loss treatment in Bolzano. Andrea is on the database, 21 and lives in Bozen.

How much information about individuals is revealed?

Information – A Measure of Surprise

The Entropy of a probability distribution **p** is:

$$H(\mathbf{p}) = -\sum \mathbf{p}(x)\log_2(\mathbf{p}(x)).$$

Example (Cover&Thomas)

Consider a Cheltenham Horse Race with 8 horses with winning chances $(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64})$. then the entropy is 2.

Label horses 0, 10, 110, 1110, 111100, 111101, 111110, 111111 then we need only only 2 bits in average to report winner.

For equally strong horses $(\frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8})$ we have an entropy of 3, and the minimal message length is also 3 bits.

Information – A Measure of Surprise

The Entropy of a probability distribution **p** is:

$$H(\mathbf{p}) = -\sum \mathbf{p}(x) \log_2(\mathbf{p}(x)).$$

Example (Cover&Thomas)

Consider a Cheltenham Horse Race with 8 horses with winning chances $(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64})$. then the entropy is 2.

Label horses $0, 10, 110, 1110, 111100, 111101, 111110, 111111$ then we need only only 2 bits in average to report winner.

For equally strong horses $(\frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8})$ we have an entropy of 3, and the minimal message length is also 3 bits.

A Priori – Surprise vs Prejudice

A father and son are on a fishing trip in the mountains of Wales. On the way back home their car has a serious accident.

The father is immediately killed and declared dead on the site of the accident. However, the son is severely injured and driven by ambulance to the next hospital.

When the son is brought into the operating theatre the surgeon exclaims "I can't do this, he is my son."

Scientific American, 1980s

A Priori – Surprise vs Prejudice

A father and son are on a fishing trip in the mountains of Wales. On the way back home their car has a serious accident.

The father is immediately killed and declared dead on the site of the accident. However, the son is severely injured and driven by ambulance to the next hospital.

When the son is brought into the operating theatre the surgeon exclaims "I can't do this, he is my son."

Scientific American, 1980s

A Priori – Surprise vs Prejudice

A father and son are on a fishing trip in the mountains of Wales. On the way back home their car has a serious accident.

The father is immediately killed and declared dead on the site of the accident. However, the son is severely injured and driven by ambulance to the next hospital.

When the son is brought into the operating theatre the surgeon exclaims "I can’t do this, he is my son."

Scientific American, 1980s

The Monty Hall Problem

- The game show proceeds as follows: First the contestant is invited to pick one of three doors (behind one is the prize) but the door is not yet opened.
- Instead, the host – legendary Monty Hall – opens one of the other doors which is empty.
- After that the contestant is given a last chance to stick with his/her door or to switch to the other closed one.
- Note that the host (knowing where the prize is) has always at least one door he can open.

The Monty Hall Problem

- The game show proceeds as follows: First the contestant is invited to pick one of three doors (behind one is the prize) but the door is not yet opened.
- Instead, the host – legendary Monty Hall – opens one of the other doors which is empty.
- After that the contestant is given a last chance to stick with his/her door or to switch to the other closed one.
- Note that the host (knowing where the prize is) has always at least one door he can open.

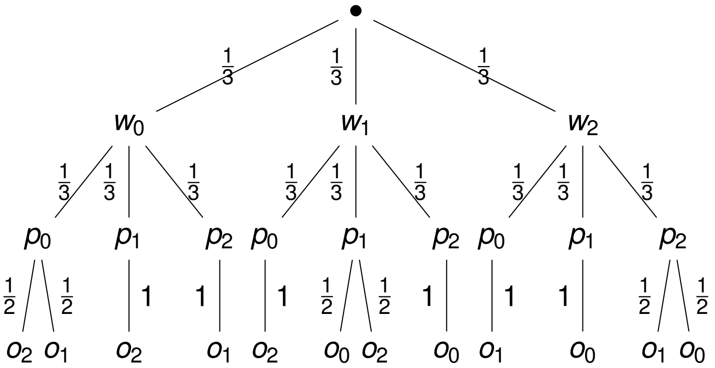
The Monty Hall Problem

- The game show proceeds as follows: First the contestant is invited to pick one of three doors (behind one is the prize) but the door is not yet opened.
- Instead, the host – legendary Monty Hall – opens one of the other doors which is empty.
- After that the contestant is given a last chance to stick with his/her door or to switch to the other closed one.
- Note that the host (knowing where the prize is) has always at least one door he can open.

The Monty Hall Problem

- The game show proceeds as follows: First the contestant is invited to pick one of three doors (behind one is the prize) but the door is not yet opened.
- Instead, the host – legendary Monty Hall – opens one of the other doors which is empty.
- After that the contestant is given a last chance to stick with his/her door or to switch to the other closed one.
- Note that the host (knowing where the prize is) has always at least one door he can open.

Optimal Strategy: To Switch or not to Switch



w_i = win behind i p_i = pick door i o_i = Monty opens door i

Certainty, Possibility, Probability

Certainty — Determinism
Model: Definite Value
e.g. $2 \in \mathbb{N}$

Possibility — Non-Determinism
Model: Set of Values
e.g. $\{2, 4, 6, 8, 10\} \in \mathcal{P}(\mathbb{N})$

Probability — Probabilistic Non-Determinism
Model: Distribution (Measure)
e.g. $(0, 0, \frac{1}{5}, 0, \frac{1}{5}, 0, \ldots) \in \mathcal{V}(\mathbb{N})$

Certainty, Possibility, Probability

Certainty — Determinism
Model: Definite Value
e.g. $2 \in \mathbb{N}$

Possibility — Non-Determinism
Model: Set of Values
e.g. $\{2, 4, 6, 8, 10\} \in \mathcal{P}(\mathbb{N})$

Probability — Probabilistic Non-Determinism
Model: Distribution (Measure)
e.g. $(0, 0, \frac{1}{5}, 0, \frac{1}{5}, 0, \ldots) \in \mathcal{V}(\mathbb{N})$

Certainty, Possibility, Probability

Certainty — Determinism
Model: Definite Value
e.g. $2 \in \mathbb{N}$

Possibility — Non-Determinism
Model: Set of Values
e.g. $\{2, 4, 6, 8, 10\} \in \mathcal{P}(\mathbb{N})$

Probability — Probabilistic Non-Determinism
Model: Distribution (Measure)
e.g. $(0, 0, \frac{1}{5}, 0, \frac{1}{5}, 0, \ldots) \in \mathcal{V}(\mathbb{N})$

Structures: Power Sets

Given a finite set (universe) Ω (of states) we can construct the power set $\mathcal{P}(\Omega)$ of Ω easily as:

$$\mathcal{P}(\Omega) = \{X \mid X \subseteq \Omega\}$$

Ordered by inclusion "$\subseteq$" this is *the* example of a lattice/order.

It can also be seen as the set of functions from S into a two element set, thus $\mathcal{P}(\Omega) = 2^\Omega$:

$$\mathcal{P}(\Omega) = \{\chi : \Omega \to \{0, 1\}\}$$

A priori, no major problems when Ω is (un)countable infinite.

Structures: Power Sets

Given a finite set (universe) Ω (of states) we can construct the power set $\mathcal{P}(\Omega)$ of Ω easily as:

$$\mathcal{P}(\Omega) = \{X \mid X \subseteq \Omega\}$$

Ordered by inclusion "$\subseteq$" this is *the* example of a lattice/order.

It can also be seen as the set of functions from S into a two element set, thus $\mathcal{P}(\Omega) = 2^{\Omega}$:

$$\mathcal{P}(\Omega) = \{\chi : \Omega \to \{0, 1\}\}$$

A priori, no major problems when Ω is (un)countable infinite.

Structures: Power Sets

Given a finite set (universe) Ω (of states) we can construct the power set $\mathcal{P}(\Omega)$ of Ω easily as:

$$\mathcal{P}(\Omega) = \{X \mid X \subseteq \Omega\}$$

Ordered by inclusion "$\subseteq$" this is *the* example of a lattice/order.

It can also be seen as the set of functions from S into a two element set, thus $\mathcal{P}(\Omega) = 2^{\Omega}$:

$$\mathcal{P}(\Omega) = \{\chi : \Omega \to \{0, 1\}\}$$

A priori, no major problems when Ω is (un)countable infinite.

Structures: Vector Spaces

Vector Spaces = Abelian Additive Group + Quantities

Given a finite set Ω we can construct the (free) vector space $\mathcal{V}(\Omega)$ of Ω as a tuple space (with $\mathbb{K}$ a field like $\mathbb{R}$ or $\mathbb{C}$):

$$\mathcal{V}(\Omega) = \{\langle \omega, x_\omega \rangle \mid \omega \in \Omega, x_\omega \in \mathbb{K}\} = \{(x_\omega)_{\omega\in\Omega} \mid x_\omega \in \mathbb{K}\}$$

As function spaces $\mathcal{V}(\Omega)$ and $\mathcal{P}(\Omega)$ are not so different:

$$\mathcal{V}(\Omega) = \{v : \Omega \to \mathbb{K}\}$$

However, there are major topological problems when Ω is (un)countable infinite.

Structures: Vector Spaces

Vector Spaces = Abelian Additive Group + Quantities

Given a finite set Ω we can construct the (free) vector space $\mathcal{V}(\Omega)$ of Ω as a tuple space (with $\mathbb{K}$ a field like $\mathbb{R}$ or $\mathbb{C}$):

$$\mathcal{V}(\Omega) = \{\langle \omega, x_\omega \rangle \mid \omega \in \Omega, x_\omega \in \mathbb{K}\} = \{(x_\omega)_{\omega \in \Omega} \mid x_\omega \in \mathbb{K}\}$$

As function spaces $\mathcal{V}(\Omega)$ and $\mathcal{P}(\Omega)$ are not so different:

$$\mathcal{V}(\Omega) = \{v : \Omega \to \mathbb{K}\}$$

However, there are major topological problems when Ω is (un)countable infinite.

Structures: Vector Spaces

Vector Spaces = Abelian Additive Group + Quantities

Given a finite set Ω we can construct the (free) vector space $\mathcal{V}(\Omega)$ of Ω as a tuple space (with $\mathbb{K}$ a field like $\mathbb{R}$ or $\mathbb{C}$):

$$\mathcal{V}(\Omega) = \{\langle \omega, x_\omega \rangle \mid \omega \in \Omega, x_\omega \in \mathbb{K}\} = \{(x_\omega)_{\omega\in\Omega} \mid x_\omega \in \mathbb{K}\}$$

As function spaces $\mathcal{V}(\Omega)$ and $\mathcal{P}(\Omega)$ are not so different:

$$\mathcal{V}(\Omega) = \{v : \Omega \to \mathbb{K}\}$$

However, there are major topological problems when Ω is (un)countable infinite.

Structures: Vector Spaces

Vector Spaces = Abelian Additive Group + Quantities

Given a finite set Ω we can construct the (free) vector space $\mathcal{V}(\Omega)$ of Ω as a tuple space (with $\mathbb{K}$ a field like $\mathbb{R}$ or $\mathbb{C}$):

$$\mathcal{V}(\Omega) = \{\langle \omega, x_\omega \rangle \mid \omega \in \Omega, x_\omega \in \mathbb{K}\} = \{(x_\omega)_{\omega \in \Omega} \mid x_\omega \in \mathbb{K}\}$$

As function spaces $\mathcal{V}(\Omega)$ and $\mathcal{P}(\Omega)$ are not so different:

$$\mathcal{V}(\Omega) = \{v : \Omega \to \mathbb{K}\}$$

However, there are major topological problems when Ω is (un)countable infinite.

Structures: Vector Spaces

Vector Spaces = Abelian Additive Group + Quantities

Given a finite set Ω we can construct the (free) vector space $\mathcal{V}(\Omega)$ of Ω as a tuple space (with $\mathbb{K}$ a field like $\mathbb{R}$ or $\mathbb{C}$):

$$\mathcal{V}(\Omega) = \{\langle \omega, x_\omega \rangle \mid \omega \in \Omega, x_\omega \in \mathbb{K}\} = \{(x_\omega)_{\omega\in\Omega} \mid x_\omega \in \mathbb{K}\}$$

As function spaces $\mathcal{V}(\Omega)$ and $\mathcal{P}(\Omega)$ are not so different:

$$\mathcal{V}(\Omega) = \{v : \Omega \to \mathbb{K}\}$$

However, there are major topological problems when Ω is (un)countable infinite.

Tuple Spaces

Theorem

All finite dimensional vector spaces are isomorphic to the (finite) Cartesian product of the underlying field $\mathbb{K}^n$ *(e.g.* $\mathbb{R}^n$ *or* $\mathbb{C}^m$*).*

Finite dimensional vectors can always be represented via their coordinates with respect to a given base, e.g.

$$\begin{aligned} x &= (x_1, x_2, x_3, \ldots, x_n) \\ y &= (y_1, y_2, y_3, \ldots, y_n) \end{aligned}$$

Algebraic Structure

$$\begin{aligned} \alpha x &= (\alpha x_1, \alpha x_2, \alpha x_3, \ldots, \alpha x_n) \\ x + y &= (x_1 + y_1, x_2 + y_2, x_3 + y_3, \ldots, x_n + y_n) \end{aligned}$$

Probability Theory

Probability theory is concerned with quantifying or measuring the chances that certain events can happen.

We consider an event space, i.e. a finite set Ω and a set $\mathcal{B} \subseteq \mathcal{P}(\Omega)$ of measurable sets in Ω which form a Boolean algebra (based on union, intersection and complement). For finite sets one can use the power-set $\mathcal{B} = \mathcal{P}(\Omega)$.

Probabilities are then assigned to event sets via a measure, i.e. a function $\Pr : \mathcal{B} \to \mathbb{R}$ or $m : \mathcal{B} \to \mathbb{R}$ or $\mu : \mathcal{B} \to \mathbb{R}$.

Note: For (uncountable) infinite Ω one needs to develop a more general measure theory [more later].

Probability Theory

Probability theory is concerned with quantifying or measuring the chances that certain events can happen.

We consider an event space, i.e. a finite set Ω and a set $\mathcal{B} \subseteq \mathcal{P}(\Omega)$ of measurable sets in Ω which form a Boolean algebra (based on union, intersection and complement). For finite sets one can use the power-set $\mathcal{B} = \mathcal{P}(\Omega)$.

Probabilities are then assigned to event sets via a measure, i.e. a function $\Pr : \mathcal{B} \rightarrow \mathbb{R}$ or $m : \mathcal{B} \rightarrow \mathbb{R}$ or $\mu : \mathcal{B} \rightarrow \mathbb{R}$.

Note: For (uncountable) infinite Ω one needs to develop a more general measure theory [more later].

Probability Theory

Probability theory is concerned with quantifying or measuring the chances that certain events can happen.

We consider an event space, i.e. a finite set Ω and a set $\mathcal{B} \subseteq \mathcal{P}(\Omega)$ of measurable sets in Ω which form a Boolean algebra (based on union, intersection and complement).
For finite sets one can use the power-set $\mathcal{B} = \mathcal{P}(\Omega)$.

Probabilities are then assigned to event sets via a measure, i.e. a function $\Pr : \mathcal{B} \to \mathbb{R}$ or $m : \mathcal{B} \to \mathbb{R}$ or $\mu : \mathcal{B} \to \mathbb{R}$.

Note: For (uncountable) infinite Ω one needs to develop a more general measure theory [more later].

Probability Theory

Probability theory is concerned with quantifying or measuring the chances that certain events can happen.

We consider an event space, i.e. a finite set Ω and a set $\mathcal{B} \subseteq \mathcal{P}(\Omega)$ of measurable sets in Ω which form a Boolean algebra (based on union, intersection and complement). For finite sets one can use the power-set $\mathcal{B} = \mathcal{P}(\Omega)$.

Probabilities are then assigned to event sets via a measure, i.e. a function $\Pr : \mathcal{B} \rightarrow \mathbb{R}$ or $m : \mathcal{B} \rightarrow \mathbb{R}$ or $\mu : \mathcal{B} \rightarrow \mathbb{R}$.

Note: For (uncountable) infinite Ω one needs to develop a more general measure theory [more later].

Probability Theory

Probability theory is concerned with quantifying or measuring the chances that certain events can happen.

We consider an event space, i.e. a finite set Ω and a set $\mathcal{B} \subseteq \mathcal{P}(\Omega)$ of measurable sets in Ω which form a Boolean algebra (based on union, intersection and complement).
For finite sets one can use the power-set $\mathcal{B} = \mathcal{P}(\Omega)$.

Probabilities are then assigned to event sets via a measure, i.e. a function $\Pr : \mathcal{B} \to \mathbb{R}$ or $m : \mathcal{B} \to \mathbb{R}$ or $\mu : \mathcal{B} \to \mathbb{R}$.

Note: For (uncountable) infinite Ω one needs to develop a more general measure theory [more later].

Probability Theory

Probability theory is concerned with quantifying or measuring the chances that certain events can happen.

We consider an event space, i.e. a finite set Ω and a set $\mathcal{B} \subseteq \mathcal{P}(\Omega)$ of measurable sets in Ω which form a Boolean algebra (based on union, intersection and complement). For finite sets one can use the power-set $\mathcal{B} = \mathcal{P}(\Omega)$.

Probabilities are then assigned to event sets via a measure, i.e. a function $\Pr : \mathcal{B} \to \mathbb{R}$ or $m : \mathcal{B} \to \mathbb{R}$ or $\mu : \mathcal{B} \to \mathbb{R}$.

Note: For (uncountable) infinite Ω one needs to develop a more general measure theory [more later].

Finite Probability Spaces

Consider a finite measurable spaces $(\Omega, \mathcal{B})$ with $|\Omega| = n$,

Definition

A probability (measure) $\Pr : \mathcal{B}$ on $(\Omega, \mathcal{B})$ has to fulfill

- $\Pr(\Omega) = 1$.
- $0 \leq \Pr(A) \leq 1$ for all $A \in \mathcal{B}$.
- $\Pr(A \cup B) = \Pr(A) + \Pr(B)$ for $A \cap B = \emptyset$.

Some further rules (which follow from the axioms above):

- $\Pr(\emptyset) = 0$,
- $\Pr(\overline{A}) = 1 - \Pr(A)$,
- $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$.
- etc.

Finite Probability Spaces

Consider a finite measurable spaces $(\Omega, \mathcal{B})$ with $|\Omega| = n$,

Definition

A probability (measure) $\Pr : \mathcal{B}$ on $(\Omega, \mathcal{B})$ has to fulfill

- $\Pr(\Omega) = 1$.
- $0 \leq \Pr(A) \leq 1$ for all $A \in \mathcal{B}$.
- $\Pr(A \cup B) = \Pr(A) + \Pr(B)$ for $A \cap B = \emptyset$.

Some further rules (which follow from the axioms above):

- $\Pr(\emptyset) = 0$,
- $\Pr(\overline{A}) = 1 - \Pr(A)$,
- $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$.
- etc.

Finite Probability Spaces

Consider a finite measurable spaces $(\Omega, \mathcal{B})$ with $|\Omega| = n$,

Definition

A probability (measure) $\Pr : \mathcal{B}$ on $(\Omega, \mathcal{B})$ has to fulfill

- $\Pr(\Omega) = 1$.
- $0 \leq \Pr(A) \leq 1$ for all $A \in \mathcal{B}$.
- $\Pr(A \cup B) = \Pr(A) + \Pr(B)$ for $A \cap B = \emptyset$.

Some further rules (which follow from the axioms above):

- $\Pr(\emptyset) = 0$,
- $\Pr(\overline{A}) = 1 - \Pr(A)$,
- $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$.
- etc.

Random Distributions

For finite probability spaces $(\Omega, \mathcal{B}, \Pr)$ with $|\Omega| = n$, we can define a probability (measure) via atoms in $\omega \in \Omega$.

Definition

A probability distribution is a function $\mathbf{p} : \Omega \to [0, 1]$, with

$$\sum_{\omega \in \Omega} \mathbf{P}(\omega) = 1.$$

If we enumerate the elements in Ω in some arbitrary way as $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ then we can also represented $\mathbf{p}$ by a (row) vector in $\mathbb{R}^n$.

$$\mathbf{p} = (\mathbf{p}(\omega_1), \mathbf{p}(\omega_2), \ldots, \mathbf{p}(\omega_n))$$

Random Distributions

For finite probability spaces $(\Omega, \mathcal{B}, \Pr)$ with $|\Omega| = n$, we can define a probability (measure) via atoms in $\omega \in \Omega$.

Definition

A probability distribution is a function $\mathbf{p} : \Omega \to [0, 1]$, with

$$\sum_{\omega \in \Omega} \mathbf{P}(\omega) = 1.$$

If we enumerate the elements in Ω in some arbitrary way as $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ then we can also represented $\mathbf{p}$ by a (row) vector in $\mathbb{R}^n$.

$$\mathbf{p} = (\mathbf{p}(\omega_1), \mathbf{p}(\omega_2), \ldots, \mathbf{p}(\omega_n))$$

Random Distributions

For finite probability spaces $(\Omega, \mathcal{B}, \Pr)$ with $|\Omega| = n$, we can define a probability (measure) via atoms in $\omega \in \Omega$.

Definition

A probability distribution is a function $\mathbf{p} : \Omega \to [0, 1]$, with

$$\sum_{\omega \in \Omega} \mathbf{P}(\omega) = 1.$$

If we enumerate the elements in Ω in some arbitrary way as $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ then we can also represented $\mathbf{p}$ by a (row) vector in $\mathbb{R}^n$.

$$\mathbf{p} = (\mathbf{p}(\omega_1), \mathbf{p}(\omega_2), \ldots, \mathbf{p}(\omega_n))$$

Random Variables

Probability distributions on a finite Ω define a probability (measure) in the obvious way

$$\Pr(A) = \sum_{\omega \in A} \mathbf{p}(\omega).$$

Definition

A random variable is function $X : \Omega \to \mathbb{R}$.

We can represent random variables as (column) vectors in $\mathbb{R}^n$.

Example

Consider a dice. The event space describes the top face of the dice, i.e. $\Omega = \{$ $\}$, define X which counts the number of eyes, e.g. $X($ $) = 5$ etc.

Random Variables

Probability distributions on a finite Ω define a probability (measure) in the obvious way

$$\Pr(A) = \sum_{\omega \in A} \mathbf{p}(\omega).$$

Definition

A random variable is function $X : \Omega \to \mathbb{R}$.

We can represent random variables as (column) vectors in $\mathbb{R}^n$.

Example

Consider a dice. The event space describes the top face of the dice, i.e. $\Omega = \{$ ⚀ , ⚁ , ⚂ , ⚃ , ⚄ , ⚅ $\}$, define X which counts the number of eyes, e.g. $X($ ⚄ $) = 5$ etc.

Random Variables

Probability distributions on a finite Ω define a probability (measure) in the obvious way

$$\Pr(A) = \sum_{\omega \in A} \mathbf{p}(\omega).$$

Definition

A random variable is function $X : \Omega \to \mathbb{R}$.

We can represent random variables as (column) vectors in $\mathbb{R}^n$.

Example

Consider a dice. The event space describes the top face of the dice, i.e. $\Omega = \{$ $\}$, define X which counts the number of eyes, e.g. $X($ $) = 5$ etc.

Random Variables

Probability distributions on a finite Ω define a probability (measure) in the obvious way

$$\Pr(A) = \sum_{\omega \in A} \mathbf{p}(\omega).$$

Definition

A random variable is function $X : \Omega \to \mathbb{R}$.

We can represent random variables as (column) vectors in $\mathbb{R}^n$.

Example

Consider a dice. The event space describes the top face of the dice, i.e. $\Omega = \{$ ⚀, ⚁, ⚂, ⚃, ⚄, ⚅ $\}$, define X which counts the number of eyes, e.g. $X($⚄$) = 5$ etc.

Moments in Probability

Expectation Value. For a random variable X and probability distribution $\mathbf{p}$ we define:

$$\mathbf{E}(X) = \sum_{\omega \in \Omega} \mathbf{p}(\omega) X(\omega) = \sum_i \mathbf{p}_i \mathbf{X}_i = \mu_X$$

One can show: $\mathbf{E}(X+Y) = \mathbf{E}(X) + \mathbf{E}(Y)$ and $\mathbf{E}(\alpha X) = \alpha \mathbf{E}(X)$.

Example

Consider a (fair) dice and previous random variable X, then

$$\mathbf{E}(X) = 1\frac{1}{6} + 2\frac{1}{6} + 3\frac{1}{6} + 4\frac{1}{6} + 5\frac{1}{6} + 6\frac{1}{6} = \frac{21}{6}$$

Variance. For random variable X and distribution $\mathbf{p}$ we define:

$$\mathrm{Var}(X) = \mathbf{E}((X - \mathbf{E}(X))^2 = \mathbf{E}(X^2) - (\mathbf{E}(X))^2.$$

The standard deviation is $\sigma_X = \sqrt{\mathrm{Var}(X)}$.

Moments in Probability

Expectation Value. For a random variable X and probability distribution $\mathbf{p}$ we define:

$$\mathbf{E}(X) = \sum_{\omega \in \Omega} \mathbf{p}(\omega) X(\omega) = \sum_i \mathbf{p}_i \mathbf{X}_i = \mu_X$$

One can show: $\mathbf{E}(X + Y) = \mathbf{E}(X) + \mathbf{E}(Y)$ and $\mathbf{E}(\alpha X) = \alpha \mathbf{E}(X)$.

Example

Consider a (fair) dice and previous random variable X, then

$$\mathbf{E}(X) = 1\frac{1}{6} + 2\frac{1}{6} + 3\frac{1}{6} + 4\frac{1}{6} + 5\frac{1}{6} + 6\frac{1}{6} = \frac{21}{6}$$

Variance. For random variable X and distribution $\mathbf{p}$ we define:

$$\mathrm{Var}(X) = \mathbf{E}((X - \mathbf{E}(X))^2 = \mathbf{E}(X^2) - (\mathbf{E}(X))^2.$$

The standard deviation is $\sigma_X = \sqrt{\mathrm{Var}(X)}$.

Moments in Probability

Expectation Value. For a random variable X and probability distribution $\mathbf{p}$ we define:

$$\mathbf{E}(X) = \sum_{\omega \in \Omega} \mathbf{p}(\omega) X(\omega) = \sum_i \mathbf{p}_i \mathbf{X}_i = \mu_X$$

One can show: $\mathbf{E}(X + Y) = \mathbf{E}(X) + \mathbf{E}(Y)$ and $\mathbf{E}(\alpha X) = \alpha \mathbf{E}(X)$.

Example

Consider a (fair) dice and previous random variable X, then

$$\mathbf{E}(X) = 1\frac{1}{6} + 2\frac{1}{6} + 3\frac{1}{6} + 4\frac{1}{6} + 5\frac{1}{6} + 6\frac{1}{6} = \frac{21}{6}$$

Variance. For random variable X and distribution $\mathbf{p}$ we define:

$$\mathrm{Var}(X) = \mathbf{E}((X - \mathbf{E}(X))^2 = \mathbf{E}(X^2) - (\mathbf{E}(X))^2.$$

The standard deviation is $\sigma_X = \sqrt{\mathrm{Var}(X)}$.

Moments in Probability

Expectation Value. For a random variable X and probability distribution $\mathbf{p}$ we define:

$$\mathbf{E}(X) = \sum_{\omega \in \Omega} \mathbf{p}(\omega) X(\omega) = \sum_i \mathbf{p}_i \mathbf{X}_i = \mu_X$$

One can show: $\mathbf{E}(X + Y) = \mathbf{E}(X) + \mathbf{E}(Y)$ and $\mathbf{E}(\alpha X) = \alpha \mathbf{E}(X)$.

Example

Consider a (fair) dice and previous random variable X, then

$$\mathbf{E}(X) = 1\frac{1}{6} + 2\frac{1}{6} + 3\frac{1}{6} + 4\frac{1}{6} + 5\frac{1}{6} + 6\frac{1}{6} = \frac{21}{6}$$

Variance. For random variable X and distribution $\mathbf{p}$ we define:

$$\mathrm{Var}(X) = \mathbf{E}((X - \mathbf{E}(X))^2 = \mathbf{E}(X^2) - (\mathbf{E}(X))^2.$$

The standard deviation is $\sigma_X = \sqrt{\mathrm{Var}(X)}$.

Bayes Theorem and Independence

Given two subsets A and B in a probability space $(\Omega, \mathcal{B}, \Pr)$. The conditional probability of A given that B has happened is

$$\Pr_B(A) = \Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

One can show Bayes Theorem which states:

$$\Pr_A(B) = \frac{\Pr_B(A)\Pr(B)}{\Pr(A)} = \frac{\Pr(A \mid B)\Pr(B)}{\Pr(A)} = \Pr(B \mid A).$$

Given two subsets A and B in a probability space $(\Omega, \mathcal{B}, \Pr)$, A and B are (probabilistically) independent if

$$\Pr(B) = \Pr(B \mid A) = \frac{\Pr(A \cap B)}{\Pr(A)} \text{ or } \Pr(A \cap B) = \Pr(A)\Pr(B).$$

Bayes Theorem and Independence

Given two subsets A and B in a probability space $(\Omega, \mathcal{B}, \Pr)$. The conditional probability of A given that B has happened is

$$\Pr_B(A) = \Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

One can show Bayes Theorem which states:

$$\Pr_A(B) = \frac{\Pr_B(A)\Pr(B)}{\Pr(A)} = \frac{\Pr(A \mid B)\Pr(B)}{\Pr(A)} = \Pr(B \mid A).$$

Given two subsets A and B in a probability space $(\Omega, \mathcal{B}, \Pr)$, A and B are (probabilistically) independent if

$$\Pr(B) = \Pr(B \mid A) = \frac{\Pr(A \cap B)}{\Pr(A)} \text{ or } \Pr(A \cap B) = \Pr(A)\Pr(B).$$

Bayes Theorem and Independence

Given two subsets A and B in a probability space $(\Omega, \mathcal{B}, \Pr)$. The conditional probability of A given that B has happened is

$$\Pr_B(A) = \Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

One can show Bayes Theorem which states:

$$\Pr_A(B) = \frac{\Pr_B(A)\Pr(B)}{\Pr(A)} = \frac{\Pr(A \mid B)\Pr(B)}{\Pr(A)} = \Pr(B \mid A).$$

Given two subsets A and B in a probability space $(\Omega, \mathcal{B}, \Pr)$, A and B are (probabilistically) independent if

$$\Pr(B) = \Pr(B \mid A) = \frac{\Pr(A \cap B)}{\Pr(A)} \quad \text{or} \quad \Pr(A \cap B) = \Pr(A)\Pr(B).$$

Bayes Theorem and Independence

Given two subsets A and B in a probability space $(\Omega, \mathcal{B}, \Pr)$. The conditional probability of A given that B has happened is

$$\Pr_B(A) = \Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

One can show Bayes Theorem which states:

$$\Pr_A(B) = \frac{\Pr_B(A)\Pr(B)}{\Pr(A)} = \frac{\Pr(A \mid B)\Pr(B)}{\Pr(A)} = \Pr(B \mid A).$$

Given two subsets A and B in a probability space $(\Omega, \mathcal{B}, \Pr)$, A and B are (probabilistically) independent if

$$\Pr(B) = \Pr(B \mid A) = \frac{\Pr(A \cap B)}{\Pr(A)} \text{ or } \Pr(A \cap B) = \Pr(A)\Pr(B).$$

Bayes Theorem and Independence

Given two subsets A and B in a probability space $(\Omega, \mathcal{B}, \Pr)$. The conditional probability of A given that B has happened is

$$\Pr_B(A) = \Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

One can show Bayes Theorem which states:

$$\Pr_A(B) = \frac{\Pr_B(A)\Pr(B)}{\Pr(A)} = \frac{\Pr(A \mid B)\Pr(B)}{\Pr(A)} = \Pr(B \mid A).$$

Given two subsets A and B in a probability space $(\Omega, \mathcal{B}, \Pr)$, A and B are (probabilistically) independent if

$$\Pr(B) = \Pr(B \mid A) = \frac{\Pr(A \cap B)}{\Pr(A)} \quad \text{or} \quad \Pr(A \cap B) = \Pr(A)\Pr(B).$$

Bayes Theorem and Independence

Given two subsets A and B in a probability space $(\Omega, \mathcal{B}, \text{Pr})$. The conditional probability of A given that B has happened is

$$\text{Pr}_B(A) = \text{Pr}(A \mid B) = \frac{\text{Pr}(A \cap B)}{\text{Pr}(B)}$$

One can show Bayes Theorem which states:

$$\text{Pr}_A(B) = \frac{\text{Pr}_B(A)\text{Pr}(B)}{\text{Pr}(A)} = \frac{\text{Pr}(A \mid B)\text{Pr}(B)}{\text{Pr}(A)} = \text{Pr}(B \mid A).$$

Given two subsets A and B in a probability space $(\Omega, \mathcal{B}, \text{Pr})$, A and B are (probabilistically) independent if

$$\text{Pr}(B) = \text{Pr}(B \mid A) = \frac{\text{Pr}(A \cap B)}{\text{Pr}(A)} \text{ or } \text{Pr}(A \cap B) = \text{Pr}(A)\text{Pr}(B).$$

Products and Probability

Given two probability spaces $(\Omega_1, \Pr_1)$ and $(\Omega_2, \Pr_2)$, to keep things simple use $\mathcal{B}_i = \mathcal{P}(\Omega_i)$. We can define a probability Pr on the cartesian product $\Omega = \Omega_1 \times \Omega_2$ via:

$$\Pr(\langle \omega_1, \omega_2 \rangle) = \Pr_1(\omega_1)\Pr_1(\omega_2)$$

If $\Pr_1$ and $\Pr_2$ correspond to probability distributions $\mathbf{p}_1$ and $\mathbf{p}_2$ and $\mathbf{p}$ to Pr then $\mathbf{p} = \mathbf{p}_1 \otimes \mathbf{p}_2$, i.e. the tensor product.

Caveat: Not all distributions on $\Omega_1 \times \Omega_2$ are a product.

Example

Consider $\Omega_1 = \{0, 1\}$ and $\Omega_2 = \{z, o\}$ and probability $\Pr(\langle 0, z \rangle) = \Pr(\langle 1, o \rangle) = \frac{1}{2}$ and $\Pr(\langle 0, o \rangle) = \Pr(\langle 1, z \rangle) = 0$ cannot be represented as a product $\mathbf{p}_1 \otimes \mathbf{p}_2$.

However, one can show: $\mathbf{p} = \frac{1}{2}(1, 0) \otimes (1, 0) + \frac{1}{2}(0, 1) \otimes (0, 1)$.

Products and Probability

Given two probability spaces $(\Omega_1, \mathrm{Pr}_1)$ and $(\Omega_2, \mathrm{Pr}_2)$, to keep things simple use $\mathcal{B}_i = \mathcal{P}(\Omega_i)$. We can define a probability Pr on the cartesian product $\Omega = \Omega_1 \times \Omega_2$ via:

$$\mathrm{Pr}(\langle \omega_1, \omega_2 \rangle) = \mathrm{Pr}_1(\omega_1)\mathrm{Pr}_1(\omega_2)$$

If Pr_1 and Pr_2 correspond to probability distributions $\mathbf{p}_1$ and $\mathbf{p}_2$ and $\mathbf{p}$ to Pr then $\mathbf{p} = \mathbf{p}_1 \otimes \mathbf{p}_2$, i.e. the tensor product.

Caveat: Not all distributions on $\Omega_1 \times \Omega_2$ are a product.

Example

Consider $\Omega_1 = \{0, 1\}$ and $\Omega_2 = \{z, o\}$ and probability $\mathrm{Pr}(\langle 0, z \rangle) = \mathrm{Pr}(\langle 1, o \rangle) = \frac{1}{2}$ and $\mathrm{Pr}(\langle 0, o \rangle) = \mathrm{Pr}(\langle 1, z \rangle) = 0$ cannot be represented as a product $\mathbf{p}_1 \otimes \mathbf{p}_2$.
However, one can show: $\mathbf{p} = \frac{1}{2}(1,0) \otimes (1,0) + \frac{1}{2}(0,1) \otimes (0,1)$.

Products and Probability

Given two probability spaces $(\Omega_1, \Pr_1)$ and $(\Omega_2, \Pr_2)$, to keep things simple use $\mathcal{B}_i = \mathcal{P}(\Omega_i)$. We can define a probability Pr on the cartesian product $\Omega = \Omega_1 \times \Omega_2$ via:

$$\Pr(\langle \omega_1, \omega_2 \rangle) = \Pr_1(\omega_1)\Pr_1(\omega_2)$$

If $\Pr_1$ and $\Pr_2$ correspond to probability distributions $\mathbf{p}_1$ and $\mathbf{p}_2$ and $\mathbf{p}$ to Pr then $\mathbf{p} = \mathbf{p}_1 \otimes \mathbf{p}_2$, i.e. the tensor product.

Caveat: Not all distributions on $\Omega_1 \times \Omega_2$ are a product.

Example

Consider $\Omega_1 = \{0, 1\}$ and $\Omega_2 = \{z, o\}$ and probability $\Pr(\langle 0, z \rangle) = \Pr(\langle 1, o \rangle) = \frac{1}{2}$ and $\Pr(\langle 0, o \rangle) = \Pr(\langle 1, z \rangle) = 0$ cannot be represented as a product $\mathbf{p}_1 \otimes \mathbf{p}_2$.

However, one can show: $\mathbf{p} = \frac{1}{2}(1, 0) \otimes (1, 0) + \frac{1}{2}(0, 1) \otimes (0, 1)$.

Products and Probability

Given two probability spaces $(\Omega_1, \Pr_1)$ and $(\Omega_2, \Pr_2)$, to keep things simple use $\mathcal{B}_i = \mathcal{P}(\Omega_i)$. We can define a probability Pr on the cartesian product $\Omega = \Omega_1 \times \Omega_2$ via:

$$\Pr(\langle \omega_1, \omega_2 \rangle) = \Pr_1(\omega_1)\Pr_1(\omega_2)$$

If $\Pr_1$ and $\Pr_2$ correspond to probability distributions $\mathbf{p}_1$ and $\mathbf{p}_2$ and $\mathbf{p}$ to Pr then $\mathbf{p} = \mathbf{p}_1 \otimes \mathbf{p}_2$, i.e. the tensor product.

Caveat: Not all distributions on $\Omega_1 \times \Omega_2$ are a product.

Example

Consider $\Omega_1 = \{0, 1\}$ and $\Omega_2 = \{z, o\}$ and probability $\Pr(\langle 0, z \rangle) = \Pr(\langle 1, o \rangle) = \frac{1}{2}$ and $\Pr(\langle 0, o \rangle) = \Pr(\langle 1, z \rangle) = 0$ cannot be represented as a product $\mathbf{p}_1 \otimes \mathbf{p}_2$.

However, one can show: $\mathbf{p} = \frac{1}{2}(1, 0) \otimes (1, 0) + \frac{1}{2}(0, 1) \otimes (0, 1)$.

Products and Probability

Given two probability spaces $(\Omega_1, \Pr_1)$ and $(\Omega_2, \Pr_2)$, to keep things simple use $\mathcal{B}_i = \mathcal{P}(\Omega_i)$. We can define a probability Pr on the cartesian product $\Omega = \Omega_1 \times \Omega_2$ via:

$$\Pr(\langle \omega_1, \omega_2 \rangle) = \Pr_1(\omega_1)\Pr_1(\omega_2)$$

If $\Pr_1$ and $\Pr_2$ correspond to probability distributions $\mathbf{p}_1$ and $\mathbf{p}_2$ and $\mathbf{p}$ to Pr then $\mathbf{p} = \mathbf{p}_1 \otimes \mathbf{p}_2$, i.e. the tensor product.

Caveat: Not all distributions on $\Omega_1 \times \Omega_2$ are a product.

Example

Consider $\Omega_1 = \{0, 1\}$ and $\Omega_2 = \{z, o\}$ and probability $\Pr(\langle 0, z \rangle) = \Pr(\langle 1, o \rangle) = \frac{1}{2}$ and $\Pr(\langle 0, o \rangle) = \Pr(\langle 1, z \rangle) = 0$ cannot be represented as a product $\mathbf{p}_1 \otimes \mathbf{p}_2$.
However, one can show: $\mathbf{p} = \frac{1}{2}(1, 0) \otimes (1, 0) + \frac{1}{2}(0, 1) \otimes (0, 1)$.

Tensor/Kronecker Product

Given a $n \times m$ matrix $\mathbf{A}$ and a $k \times l$ matrix $\mathbf{B}$:

$$\mathbf{A} = \begin{pmatrix} a_{11} & \dots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nm} \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} b_{11} & \dots & b_{1l} \\ \vdots & \ddots & \vdots \\ b_{k1} & \dots & b_{kl} \end{pmatrix}$$

The tensor or Kronecker product $\mathbf{A} \otimes \mathbf{B}$ is a $nk \times ml$ matrix:

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & \dots & a_{1m}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{n1}\mathbf{B} & \dots & a_{nm}\mathbf{B} \end{pmatrix}$$

Special cases are square matrices ($n = m$ and $k = l$) and vectors (row $n = k = 1$, column $m = l = 1$).

Tensor/Kronecker Product

Given a $n \times m$ matrix **A** and a $k \times l$ matrix **B**:

$$\mathbf{A} = \begin{pmatrix} a_{11} & \dots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nm} \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} b_{11} & \dots & b_{1l} \\ \vdots & \ddots & \vdots \\ b_{k1} & \dots & b_{kl} \end{pmatrix}$$

The tensor or Kronecker product $\mathbf{A} \otimes \mathbf{B}$ is a $nk \times ml$ matrix:

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & \dots & a_{1m}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{n1}\mathbf{B} & \dots & a_{nm}\mathbf{B} \end{pmatrix}$$

Special cases are square matrices ($n = m$ and $k = l$) and vectors (row $n = k = 1$, column $m = l = 1$).

Tensor/Kronecker Product

Given a $n \times m$ matrix $\mathbf{A}$ and a $k \times l$ matrix $\mathbf{B}$:

$$\mathbf{A} = \begin{pmatrix} a_{11} & \dots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nm} \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} b_{11} & \dots & b_{1l} \\ \vdots & \ddots & \vdots \\ b_{k1} & \dots & b_{kl} \end{pmatrix}$$

The tensor or Kronecker product $\mathbf{A} \otimes \mathbf{B}$ is a $nk \times ml$ matrix:

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & \dots & a_{1m}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{n1}\mathbf{B} & \dots & a_{nm}\mathbf{B} \end{pmatrix}$$

Special cases are square matrices ($n = m$ and $k = l$) and vectors (row $n = k = 1$, column $m = l = 1$).

Correlation

The covariance of two random variables X and Y is:

$$\mathrm{Cov}(X, Y) = \mathbf{E}((X - \mathbf{E}(X))\mathbf{E}((Y - \mathbf{E}(Y)) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)$$

The correlation coefficient is $\rho(X, Y) = \mathrm{Cov}(X, Y)/(\sigma_X \sigma_Y)$.

For independent random variables X and Y – i.e. if we have $\Pr((X = x_j) \cap (Y = y_k)) = \Pr(X = x_j)\Pr(Y = y_k)$:

$$\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y) \text{ and } \mathrm{Cov}(X, Y) = \rho(X, Y) = 0.$$

Note that $\rho(X, Y) = 0$ does **not** imply independence.

Example

$\Pr(X = x) = \frac{1}{3}$ for $x = -1, 0, 1$ and $Y = X^2$, then $\rho(X, Y) = 0$.

Correlation

The covariance of two random variables X and Y is:

$$\mathrm{Cov}(X, Y) = \mathbf{E}((X - \mathbf{E}(X))\mathbf{E}((Y - \mathbf{E}(Y)) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)$$

The correlation coefficient is $\rho(X, Y) = \mathrm{Cov}(X, Y)/(\sigma_X \sigma_Y)$.

For independent random variables X and Y – i.e. if we have $\Pr((X = x_j) \cap (Y = y_k)) = \Pr(X = x_j)\Pr(Y = y_k)$:

$$\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y) \text{ and } \mathrm{Cov}(X, Y) = \rho(X, Y) = 0.$$

Note that $\rho(X, Y) = 0$ does **not** imply independence.

Example

$\Pr(X = x) = \frac{1}{3}$ for $x = -1, 0, 1$ and $Y = X^2$, then $\rho(X, Y) = 0$.

Correlation

The covariance of two random variables X and Y is:

$$\mathrm{Cov}(X,Y) = \mathbf{E}((X - \mathbf{E}(X))\mathbf{E}((Y - \mathbf{E}(Y)) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)$$

The correlation coefficient is $\rho(X,Y) = \mathrm{Cov}(X,Y)/(\sigma_X\sigma_Y)$.

For independent random variables X and Y – i.e. if we have $\Pr((X = x_j) \cap (Y = y_k)) = \Pr(X = x_j)\Pr(Y = y_k)$:

$$\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y) \text{ and } \mathrm{Cov}(X,Y) = \rho(X,Y) = 0.$$

Note that $\rho(X,Y) = 0$ does **not** imply independence.

Example

$\Pr(X = x) = \frac{1}{3}$ for $x = -1, 0, 1$ and $Y = X^2$, then $\rho(X,Y) = 0$.

Random or Stochastic Processes

Definition

A random process (or stochastic process) $\{X_t \mid t \in T\}$ is a sequences of random variables X_i.

Depending on the kind of 'time' (usually a group or semi-group) one can distinguish between discrete time processes (with $T = \mathbb{Z}$ or $\mathbb{T} = \mathbb{N}$), and continuous time processes (with $T = \mathbb{R}$).

Typically, one can ask, for example, that all X_i in a random process are identically distributed and independent, i.i.d, e.g. coin flips or rolling dices.

One can also allow that the X_i's depend on all or some previous X_j, a particular case are Markov processes or chains [more tomorrow], e.g. random walks.

Random or Stochastic Processes

Definition

A random process (or stochastic process) $\{X_t \mid t \in T\}$ is a sequences of random variables X_i.

Depending on the kind of 'time' (usually a group or semi-group) one can distinguish between discrete time processes (with $T = \mathbb{Z}$ or $\mathbb{T} = \mathbb{N}$), and continuous time processes (with $T = \mathbb{R}$).

Typically, one can ask, for example, that all X_i in a random process are identically distributed and independent, i.i.d, e.g. coin flips or rolling dices.

One can also allow that the X_i's depend on all or some previous X_j, a particular case are Markov processes or chains [more tomorrow], e.g. random walks.

Random or Stochastic Processes

Definition

A random process (or stochastic process) $\{X_t \mid t \in T\}$ is a sequences of random variables X_i.

Depending on the kind of 'time' (usually a group or semi-group) one can distinguish between discrete time processes (with $T = \mathbb{Z}$ or $\mathbb{T} = \mathbb{N}$), and continuous time processes (with $T = \mathbb{R}$).

Typically, one can ask, for example, that all X_i in a random process are identically distributed and independent, i.i.d, e.g. coin flips or rolling dices.

One can also allow that the X_i's depend on all or some previous X_j, a particular case are Markov processes or chains [more tomorrow], e.g. random walks.

Random or Stochastic Processes

Definition

A random process (or stochastic process) $\{X_t \mid t \in T\}$ is a sequences of random variables X_i.

Depending on the kind of 'time' (usually a group or semi-group) one can distinguish between discrete time processes (with $T = \mathbb{Z}$ or $\mathbb{T} = \mathbb{N}$), and continuous time processes (with $T = \mathbb{R}$).

Typically, one can ask, for example, that all X_i in a random process are identically distributed and independent, i.i.d, e.g. coin flips or rolling dices.

One can also allow that the X_i's depend on all or some previous X_j, a particular case are Markov processes or chains [more tomorrow], e.g. random walks.

Discrete Time Markov Chain

Given a finite set of states $\Omega = \{s_1, \ldots, s_r\}$.

A discrete time Markov chain (DTMC) on Ω is defined via a stochastic matrix $\mathbf{P}$ as a above, i.e. an $r \times r$ (square) matrix with entries $0 \leq p_{ij} \leq 1$ and such that all row sums are equal to one, i.e.

$$\sum_j p_{ij} = 1.$$

The entry p_{ij} gives the conditional probability that from state s_i we go to state s_j in one descrete) time step, i.e. $T = \mathbb{Z}$ or $\mathbb{N}$. That is

$$p_{ij} = \Pr(X_{n+1} = s_j \mid X_n = s_i)$$

which is independent of n and also considers only the next/previous time step (memory-less property).

Discrete Time Markov Chain

Given a finite set of states $\Omega = \{s_1, \ldots, s_r\}$.

A discrete time Markov chain (DTMC) on Ω is defined via a stochastic matrix **P** as a above, i.e. an $r \times r$ (square) matrix with entries $0 \leq p_{ij} \leq 1$ and such that all row sums are equal to one, i.e.

$$\sum_j p_{ij} = 1.$$

The entry p_{ij} gives the conditional probability that from state s_i we go to state s_j in one descrete) time step, i.e. $T = \mathbb{Z}$ or $\mathbb{N}$. That is

$$p_{ij} = \Pr(X_{n+1} = s_j \mid X_n = s_i)$$

which is independent of n and also considers only the next/previous time step (memory-less property).

Discrete Time Markov Chain

Given a finite set of states $\Omega = \{s_1, \ldots, s_r\}$.

A discrete time Markov chain (DTMC) on Ω is defined via a stochastic matrix $\mathbf{P}$ as a above, i.e. an $r \times r$ (square) matrix with entries $0 \leq p_{ij} \leq 1$ and such that all row sums are equal to one, i.e.

$$\sum_j p_{ij} = 1.$$

The entry p_{ij} gives the conditional probability that from state s_i we go to state s_j in one descrete) time step, i.e. $T = \mathbb{Z}$ or $\mathbb{N}$. That is

$$p_{ij} = \Pr(X_{n+1} = s_j \mid X_n = s_i)$$

which is independent of n and also considers only the next/previous time step (memory-less property).

Discrete Time Markov Processes

Let **P** be the transition matrix of a DTMC. The entry in $p_{ij}^{(n)}$ in the n-th matrix power $\mathbf{P}^n$ gives the probability that the Markov chain, starting in state s_i, will be in state s_j after exactly n steps.

At any time step we can describe the probabilities of being in a certain state s_i by a probability u_i. These probabilities define a probability distribution, i.e. a row vector

$$\mathbf{u} = (u_1, u_2, \cdots, u_r)$$

such that $0 \leq u_i \leq 1$ and $\sum_i u_i = 1$.

For any stochastic matrix **P** and probability distribution **u** the multiplication **uP** is again a probability distribution.

Discrete Time Markov Processes

Let **P** be the transition matrix of a DTMC. The entry in $p_{ij}^{(n)}$ in the n-th matrix power $\mathbf{P}^n$ gives the probability that the Markov chain, starting in state s_i, will be in state s_j after exactly n steps.

At any time step we can describe the probabilities of being in a certain state s_i by a probability u_i. These probabilities define a probability distribution, i.e. a row vector

$$\mathbf{u} = (u_1, u_2, \cdots, u_r)$$

such that $0 \leq u_i \leq 1$ and $\sum_i u_i = 1$.

For any stochastic matrix **P** and probability distribution **u** the multiplication **uP** is again a probability distribution.

Discrete Time Markov Processes

Let **P** be the transition matrix of a DTMC. The entry in $p_{ij}^{(n)}$ in the n-th matrix power $\mathbf{P}^n$ gives the probability that the Markov chain, starting in state s_i, will be in state s_j after exactly n steps.

At any time step we can describe the probabilities of being in a certain state s_i by a probability u_i. These probabilities define a probability distribution, i.e. a row vector

$$\mathbf{u} = (u_1, u_2, \cdots, u_r)$$

such that $0 \leq u_i \leq 1$ and $\sum_i u_i = 1$.

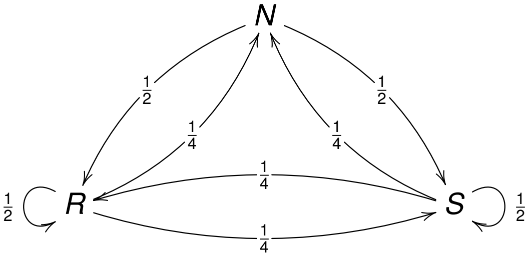
For any stochastic matrix **P** and probability distribution **u** the multiplication **uP** is again a probability distribution.

The Land of Oz

The Land of Oz is blessed with many things, but not by good weather. They never have two nice days in a row. If they have a nice day, the chance of rain or snow the next day are the same. If there is rain or snow the chances are even that the weather stays the same for the next day. If there is a change from snow or rain, only half of the time is this a change to a nice day.

The Land of Oz

From this we obtain the transition probabilities between nice (N), rainy (R) and snowy (S) days:



The Land of Oz

We can then define the following transition matrix:

$$\mathbf{P} = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix}$$

From Grinstead & Snell: *Introduction to Probability*, p406; available as GNU book on http://www.dartmouth.edu/~chance

The Land of Oz

We can then define the following transition matrix:

$$\mathbf{P} = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix}$$

From Grinstead & Snell: *Introduction to Probability*, p406;
available as GNU book on http://www.dartmouth.edu/∼chance

The Land of Oz

Consider the initial probability distributions $\mathbf{u} = (0, 1, 0)$ and $\mathbf{v} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ in the Oz Example.

The vector **u** describes a situation where we are certain that we start with a nice day (N), while **v** corresponds to one where we assume the same chances of having a rainy (R), nice (N) or snowy (S) day.

The Land of Oz

Consider the initial probability distributions $\mathbf{u} = (0, 1, 0)$ and $\mathbf{v} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ in the Oz Example.

$$\mathbf{uP} = \left(\frac{1}{2}, 0, \frac{1}{2}\right) \quad \mathbf{uP}^2 = \left(\frac{3}{8}, \frac{1}{4}, \frac{3}{8}\right) \quad \ldots$$

The Land of Oz

Consider the initial probability distributions $\mathbf{u} = (0, 1, 0)$ and $\mathbf{v} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ in the Oz Example.

$$
\begin{aligned}
\mathbf{vP}^0 &= (0.33333, 0.33333, 0.33333) \\
\mathbf{vP}^1 &= (0.41667, 0.16667, 0.41667) \\
\mathbf{vP}^2 &= (0.39583, 0.20833, 0.39583) \\
\mathbf{vP}^3 &= (0.40104, 0.19792, 0.40104) \\
\mathbf{vP}^4 &= (0.39974, 0.20052, 0.39974) \\
\ldots \\
\mathbf{vP}^{100} &= (0.40000, 0.20000, 0.40000)
\end{aligned}
$$

Convention

Note that in the theory of Markov chains one usually is concerned with probability distributions as row vectors. Therefore, probability vectors are post-multiplied by the stochastic matrix **P** defining a Markov chain.

The usual pre-multiplication could be realised via:

$$\mathbf{P}\mathbf{u} = (\mathbf{u}^t\mathbf{P}^T)^t$$

Measure Theory = Infinite Probability

If we have infinite (countable or uncountable) "universes" Ω then there are a number of problems one has to resolve when we want to define probabilities $\Pr(A)$ or measures $\mu(A)$ for $A \subseteq \Omega$.

E.g. consider the real interval $[0, 1]$; it is impossible to have

1. $\mu(x) > 0$ for all $x \in [0, 1]$, or
2. $\mu(x) = \mu(y)$ for all $x, y \in [0, 1]$, and
3. $\mu([0, 1]) = 1 < \infty$

Similarly, if we consider infinite sequences of events, e.g. coin flips, then the probability of any particular sequence is zero:

$$\prod_{i=0}^{\infty} \frac{1}{2} = \lim_{i \to \infty} \left(\frac{1}{2}\right)^n = 0$$

Avoid dealing with expressions like $\sum_{i=0}^{\infty} p_i$ and $\prod_{i=0}^{\infty} p_i$.

Measure Theory = Infinite Probability

If we have infinite (countable or uncountable) "universes" Ω then there are a number of problems one has to resolve when we want to define probabilities $\Pr(A)$ or measures $\mu(A)$ for $A \subseteq \Omega$.

E.g. consider the real interval $[0, 1]$; it is impossible to have

1. $\mu(x) > 0$ for all $x \in [0, 1]$, or
2. $\mu(x) = \mu(y)$ for all $x, y \in [0, 1]$, and
3. $\mu([0, 1]) = 1 < \infty$

Similarly, if we consider infinite sequences of events, e.g. coin flips, then the probability of any particular sequence is zero:

$$\prod_{i=0}^{\infty} \frac{1}{2} = \lim_{i \to \infty} \left(\frac{1}{2}\right)^n = 0$$

Avoid dealing with expressions like $\sum_{i=0}^{\infty} p_i$ and $\prod_{i=0}^{\infty} p_i$.

Measure Theory = Infinite Probability

If we have infinite (countable or uncountable) "universes" Ω then there are a number of problems one has to resolve when we want to define probabilities $\Pr(A)$ or measures $\mu(A)$ for $A \subseteq \Omega$.

E.g. consider the real interval $[0,1]$; it is impossible to have

1. $\mu(x) > 0$ for all $x \in [0,1]$, or
2. $\mu(x) = \mu(y)$ for all $x, y \in [0,1]$, and
3. $\mu([0,1]) = 1 < \infty$

Similarly, if we consider infinite sequences of events, e.g. coin flips, then the probability of any particular sequence is zero:

$$\prod_{i=0}^{\infty} \frac{1}{2} = \lim_{i \to \infty} \left(\frac{1}{2}\right)^n = 0$$

Avoid dealing with expressions like $\sum_{i=0}^{\infty} p_i$ and $\prod_{i=0}^{\infty} p_i$.

Measure Theory = Infinite Probability

If we have infinite (countable or uncountable) "universes" Ω then there are a number of problems one has to resolve when we want to define probabilities $\Pr(A)$ or measures $\mu(A)$ for $A \subseteq \Omega$.

E.g. consider the real interval $[0, 1]$; it is impossible to have

1. $\mu(x) > 0$ for all $x \in [0, 1]$, or
2. $\mu(x) = \mu(y)$ for all $x, y \in [0, 1]$, and
3. $\mu([0, 1]) = 1 < \infty$

Similarly, if we consider infinite sequences of events, e.g. coin flips, then the probability of any particular sequence is zero:

$$\prod_{i=0}^{\infty} \frac{1}{2} = \lim_{i \to \infty} \left(\frac{1}{2}\right)^n = 0$$

Avoid dealing with expressions like $\sum_{i=0}^{\infty} p_i$ and $\prod_{i=0}^{\infty} p_i$.

Measure Theory = Infinite Probability

If we have infinite (countable or uncountable) "universes" Ω then there are a number of problems one has to resolve when we want to define probabilities $\Pr(A)$ or measures $\mu(A)$ for $A \subseteq \Omega$.

E.g. consider the real interval $[0,1]$; it is impossible to have

1. $\mu(x) > 0$ for all $x \in [0,1]$, or
2. $\mu(x) = \mu(y)$ for all $x, y \in [0,1]$, and
3. $\mu([0,1]) = 1 < \infty$

Similarly, if we consider infinite sequences of events, e.g. coin flips, then the probability of any particular sequence is zero:

$$\prod_{i=0}^{\infty} \frac{1}{2} = \lim_{i \to \infty} \left(\frac{1}{2}\right)^n = 0$$

Avoid dealing with expressions like $\sum_{i=0}^{\infty} p_i$ and $\prod_{i=0}^{\infty} p_i$.

Measure Theory = Infinite Probability

If we have infinite (countable or uncountable) "universes" Ω then there are a number of problems one has to resolve when we want to define probabilities $\Pr(A)$ or measures $\mu(A)$ for $A \subseteq \Omega$.

E.g. consider the real interval $[0,1]$; it is impossible to have

1. $\mu(x) > 0$ for all $x \in [0,1]$, or
2. $\mu(x) = \mu(y)$ for all $x, y \in [0,1]$, and
3. $\mu([0,1]) = 1 < \infty$

Similarly, if we consider infinite sequences of events, e.g. coin flips, then the probability of any particular sequence is zero:

$$\prod_{i=0}^{\infty} \frac{1}{2} = \lim_{i \to \infty} \left(\frac{1}{2}\right)^n = 0$$

Avoid dealing with expressions like $\sum_{i=0}^{\infty} p_i$ and $\prod_{i=0}^{\infty} p_i$.

Measure Theory = Infinite Probability

If we have infinite (countable or uncountable) "universes" Ω then there are a number of problems one has to resolve when we want to define probabilities $\Pr(A)$ or measures $\mu(A)$ for $A \subseteq \Omega$.

E.g. consider the real interval $[0, 1]$; it is impossible to have

1. $\mu(x) > 0$ for all $x \in [0, 1]$, or
2. $\mu(x) = \mu(y)$ for all $x, y \in [0, 1]$, and
3. $\mu([0, 1]) = 1 < \infty$

Similarly, if we consider infinite sequences of events, e.g. coin flips, then the probability of any particular sequence is zero:

$$\prod_{i=0}^{\infty} \frac{1}{2} = \lim_{i \to \infty} \left(\frac{1}{2}\right)^n = 0$$

Avoid dealing with expressions like $\sum_{i=0}^{\infty} p_i$ and $\prod_{i=0}^{\infty} p_i$.

Measurable Spaces

Definition

Given any set Ω. A family σ of sub-sets $\sigma \subseteq \mathcal{P}(\Omega)$ is called a σ-algebra iff

1. $\emptyset \in \sigma$ and $\Omega \in \sigma$.
2. $\bigcap_{i=0}^{\infty} S_i \in \sigma$ for $S_i \in \sigma$ (countable).
3. $\Omega \setminus S \in \sigma$ for $S \in \sigma$.

We say that (Ω, σ) is a measurable space, and $S \in \sigma$ are measurable sets.

By de Morgan we have also: $\bigcup_{i=0}^{\infty} S_i \in \sigma$ for $S_i \in \sigma$ (countable).

Measurable Spaces

Definition

Given any set Ω. A family σ of sub-sets $\sigma \subseteq \mathcal{P}(\Omega)$ is called a σ-algebra iff

1. $\emptyset \in \sigma$ and $\Omega \in \sigma$.
2. $\bigcap_{i=0}^{\infty} S_i \in \sigma$ for $S_i \in \sigma$ (countable).
3. $\Omega \setminus S \in \sigma$ for $S \in \sigma$.

We say that (Ω, σ) is a measurable space, and $S \in \sigma$ are measurable sets.

By de Morgan we have also: $\bigcup_{i=0}^{\infty} S_i \in \sigma$ for $S_i \in \sigma$ (countable).

Measures and Measurable Functions

Definition

Given a measurable space (Ω, σ) then $\mu : \sigma \to \mathbb{R}^+$ is a (finite) measure if

1. $\mu(\emptyset) = 0$ (for $\mu(\Omega) = 1$ we have a probability measure).
2. $\mu(\bigcup_{i=0}^{\infty} S_i) = \sum_{i=0}^{\infty} \mu(S_i)$ for $S_i \in \sigma$ with $S_i \cap S_j = \emptyset$ for $i \neq j$.

Definition

A function $f : \Omega \to \Omega'$ between two measure spaces spaces (Ω, σ, μ) and (Ω', τ', μ') is called

measurable iff $\forall S \in \sigma' : f^{-1}(S) \in \sigma$.

measure preserving iff $\forall S \in \sigma'$ also $\mu'(S') = \mu(f^{-1}(S))$.

Measures and Measurable Functions

Definition

Given a measurable space (Ω, σ) then $\mu : \sigma \rightarrow \mathbb{R}^+$ is a (finite) measure if

1. $\mu(\emptyset) = 0$ (for $\mu(\Omega) = 1$ we have a probability measure).
2. $\mu(\bigcup_{i=0}^{\infty} S_i) = \sum_{i=0}^{\infty} \mu(S_i)$ for $S_i \in \sigma$ with $S_i \cap S_j = \emptyset$ for $i \neq j$.

Definition

A function $f : \Omega \rightarrow \Omega'$ between two measure spaces spaces (Ω, σ, μ) and (Ω', τ', μ') is called

measurable iff $\forall S \in \sigma' : \; f^{-1}(S) \in \sigma$.

measure preserving iff $\forall S \in \sigma'$ also $\mu'(S') = \mu(f^{-1}(S))$.

Toplological Spaces

Definition

A topological space is a set Ω together with a family of sub-sets $\tau \subseteq \mathcal{P}(\Omega)$, the topology (of open sets), iff

1. $\emptyset \in \tau$ and $\Omega \in \tau$.
2. $\bigcap_{i=0}^{n} O_i \in \tau$ for $O_i \in \tau$ (finite).
3. $\bigcup_{i \in I} O_i \in \tau$ for $O_i \in \tau$ (arbritrary).

The sets $O \in \tau$ are called open sets. The complements $A = O \setminus O$ of open sets are closed sets.

One can also define a topology in other ways, e.g. starting with closed sets (in which case one has finite unions and arbitrary intersections).

Toplological Spaces

Definition

A topological space is a set Ω together with a family of sub-sets $\tau \subseteq \mathcal{P}(\Omega)$, the topology (of open sets), iff

1. $\emptyset \in \tau$ and $\Omega \in \tau$.
2. $\bigcap_{i=0}^{n} O_i \in \tau$ for $O_i \in \tau$ (finite).
3. $\bigcup_{i \in I} O_i \in \tau$ for $O_i \in \tau$ (arbritrary).

The sets $O \in \tau$ are called open sets. The complements $A = O \setminus O$ of open sets are closed sets.

One can also define a topology in other ways, e.g. starting with closed sets (in which case one has finite unions and arbitrary intersections).

Generating Measure Spaces

We can always construct a measure space from a base set $\mathcal{B} \subseteq \mathcal{P}(\Omega)$ (not necessarily from singletons or atoms) and an appropriate measure μ defined on $\mathcal{B}$.

- Generate a unique σ-algebra from $\mathcal{B}$ via complements and countable intersections/unions from sets in $\mathcal{B}$.
- The function $\mu : \mathcal{B} \to \mathbb{R}$ can be extended to this σ-algebra in the obvious way (e.g. $\mu(\Omega \setminus B) = 1 - \mu(B)$ etc.)

Example

The Lebesgue measure on $[0, 1]$ is defined via the base $\mathcal{B} = \{[a, b] \mid a, b \in [0, 1]\}$, i.e. all sub-intervals, with $\mu([a, b]) = b - a$ (also base for the standard topology, i.e. Borel measure).

One can use the Axiom of Choice to construct non-measurable sets $X \subseteq [0, 1]$, e.g. Vitali sets, Banach-Tarski paradox

Generating Measure Spaces

We can always construct a measure space from a base set $\mathcal{B} \subseteq \mathcal{P}(\Omega)$ (not necessarily from singletons or atoms) and an appropriate measure μ defined on $\mathcal{B}$.

- Generate a unique σ-algebra from $\mathcal{B}$ via complements and countable intersections/unions from sets in $\mathcal{B}$.
- The function $\mu : \mathcal{B} \rightarrow \mathbb{R}$ can be extended to this σ-algebra in the obvious way (e.g. $\mu(\Omega \setminus B) = 1 - \mu(B)$ etc.)

Example

The Lebesgue measure on $[0, 1]$ is defined via the base $\mathcal{B} = \{[a, b] \mid a, b \in [0, 1]\}$, i.e. all sub-intervals, with $\mu([a, b]) = b - a$ (also base for the standard topology, i.e. Borel measure).

One can use the Axiom of Choice to construct non-measurable sets $X \subseteq [0, 1]$, e.g. Vitali sets, Banach-Tarski paradox

Generating Measure Spaces

We can always construct a measure space from a base set $\mathcal{B} \subseteq \mathcal{P}(\Omega)$ (not necessarily from singletons or atoms) and an appropriate measure μ defined on $\mathcal{B}$.

- Generate a unique σ-algebra from $\mathcal{B}$ via complements and countable intersections/unions from sets in $\mathcal{B}$.
- The function $\mu : \mathcal{B} \to \mathbb{R}$ can be extended to this σ-algebra in the obvious way (e.g. $\mu(\Omega \setminus B) = 1 - \mu(B)$ etc.)

Example

The Lebesgue measure on $[0, 1]$ is defined via the base $\mathcal{B} = \{[a, b] \mid a, b \in [0, 1]\}$, i.e. all sub-intervals, with $\mu([a, b]) = b - a$ (also base for the standard topology, i.e. Borel measure).

One can use the Axiom of Choice to construct non-measurable sets $X \subseteq [0, 1]$, e.g. Vitali sets, Banach-Tarski paradox

Generating Measure Spaces

We can always construct a measure space from a base set $\mathcal{B} \subseteq \mathcal{P}(\Omega)$ (not necessarily from singletons or atoms) and an appropriate measure μ defined on $\mathcal{B}$.

- Generate a unique σ-algebra from $\mathcal{B}$ via complements and countable intersections/unions from sets in $\mathcal{B}$.
- The function $\mu : \mathcal{B} \to \mathbb{R}$ can be extended to this σ-algebra in the obvious way (e.g. $\mu(\Omega \setminus B) = 1 - \mu(B)$ etc.)

Example

The Lebesgue measure on $[0, 1]$ is defined via the base $\mathcal{B} = \{[a, b] \mid a, b \in [0, 1]\}$, i.e. all sub-intervals, with $\mu([a, b]) = b - a$ (also base for the standard topology, i.e. Borel measure).

One can use the Axiom of Choice to construct non-measurable sets $X \subseteq [0, 1]$, e.g. Vitali sets, Banach-Tarski paradox

Generating Measure Spaces

We can always construct a measure space from a base set $\mathcal{B} \subseteq \mathcal{P}(\Omega)$ (not necessarily from singletons or atoms) and an appropriate measure μ defined on $\mathcal{B}$.

- Generate a unique σ-algebra from $\mathcal{B}$ via complements and countable intersections/unions from sets in $\mathcal{B}$.
- The function $\mu : \mathcal{B} \to \mathbb{R}$ can be extended to this σ-algebra in the obvious way (e.g. $\mu(\Omega \setminus B) = 1 - \mu(B)$ etc.)

Example

The Lebesgue measure on $[0, 1]$ is defined via the base $\mathcal{B} = \{[a, b] \mid a, b \in [0, 1]\}$, i.e. all sub-intervals, with $\mu([a, b]) = b - a$ (also base for the standard topology, i.e. Borel measure).

One can use the Axiom of Choice to construct non-measurable sets $X \subseteq [0, 1]$, e.g. Vitali sets, Banach-Tarski paradox

Measures on Trace Spaces

The set of infinite paths on $\{0, 1\}$ is (uncountable) infinite; every $0/1$ sequence is the binary representation of a real in $[0, 1]$.

We need to define a measure structure, i.e. σ-algebra, on this space. This can be done as before by considering as base $\mathcal{B}$.

Definition

Given a (finite) set of states S. A cylinder set of a finite path $\pi = s_0 \ldots s_n$ with $s_i \in S$ is the set of all paths $s_0 \ldots s_n \ldots$.

For infinite paths take the σ-algebra generated by all cylinders with probability $\Pr(s_0 \ldots s_n)$ to define a measure space, cf. Billingsly, Baier & Katoen: Principles of Model Checking.

Caveat: Not all trace sets are generated from cylinder sets, i.e. there are non-measurable trace properties.

Measures on Trace Spaces

The set of infinite paths on $\{0, 1\}$ is (uncountable) infinite; every $0/1$ sequence is the binary representation of a real in $[0, 1]$.

We need to define a measure structure, i.e. σ-algebra, on this space. This can be done as before by considering as base $\mathcal{B}$.

Definition

Given a (finite) set of states S. A cylinder set of a finite path $\pi = s_0 \dots s_n$ with $s_i \in S$ is the set of all paths $s_0 \dots s_n \dots$.

For infinite paths take the σ-algebra generated by all cylinders with probability $\Pr(s_0 \dots s_n)$ to define a measure space, cf. Billingsly, Baier & Katoen: Principles of Model Checking.

Caveat: Not all trace sets are generated from cylinder sets, i.e. there are non-measurable trace properties.

Measures on Trace Spaces

The set of infinite paths on $\{0, 1\}$ is (uncountable) infinite; every $0/1$ sequence is the binary representation of a real in $[0, 1]$.

We need to define a measure structure, i.e. σ-algebra, on this space. This can be done as before by considering as base $\mathcal{B}$.

Definition

Given a (finite) set of states S. A cylinder set of a finite path $\pi = s_0 \ldots s_n$ with $s_i \in S$ is the set of all paths $s_0 \ldots s_n \ldots$.

For infinite paths take the σ-algebra generated by all cylinders with probability $\Pr(s_0 \ldots s_n)$ to define a measure space, cf. Billingsly, Baier & Katoen: Principles of Model Checking.

Caveat: Not all trace sets are generated from cylinder sets, i.e. there are non-measurable trace properties.

Measures on Trace Spaces

The set of infinite paths on $\{0, 1\}$ is (uncountable) infinite; every $0/1$ sequence is the binary representation of a real in $[0, 1]$.

We need to define a measure structure, i.e. σ-algebra, on this space. This can be done as before by considering as base $\mathcal{B}$.

Definition

Given a (finite) set of states S. A cylinder set of a finite path $\pi = s_0 \ldots s_n$ with $s_i \in S$ is the set of all paths $s_0 \ldots s_n \ldots$.

For infinite paths take the σ-algebra generated by all cylinders with probability $\Pr(s_0 \ldots s_n)$ to define a measure space, cf. Billingsly, Baier & Katoen: Principles of Model Checking.

Caveat: Not all trace sets are generated from cylinder sets, i.e. there are non-measurable trace properties.

Measures on Trace Spaces

The set of infinite paths on $\{0, 1\}$ is (uncountable) infinite; every $0/1$ sequence is the binary representation of a real in $[0, 1]$.

We need to define a measure structure, i.e. σ-algebra, on this space. This can be done as before by considering as base $\mathcal{B}$.

Definition

Given a (finite) set of states S. A cylinder set of a finite path $\pi = s_0 \ldots s_n$ with $s_i \in S$ is the set of all paths $s_0 \ldots s_n \ldots$.

For infinite paths take the σ-algebra generated by all cylinders with probability $\Pr(s_0 \ldots s_n)$ to define a measure space, cf. Billingsly, Baier & Katoen: Principles of Model Checking.

Caveat: Not all trace sets are generated from cylinder sets, i.e. there are non-measurable trace properties.

Integrals

For a measure space (Ω, σ, μ) one can define integral(s) for random variables, i.e. functions X or f, on Ω:

$$\mathbf{E}(f) = \int_\Omega f(\omega) d\mu(\omega)$$

Typically one starts with step functions $t \in \mathcal{T}$ with $t : \Omega \rightarrow \mathbb{R}$ which are constant on some base sets in $\mathcal{B}$ or the σ-algebra σ.

Example (Step functions on [0, 1])

For $t = \sum_i t_i$ with $t_i(\omega) = c_i \in \mathbb{R}$ for ω in interval I_i s.t. $I_i = [a_i, b_i]$ and $\bigcup_i I_i = [0, 1]$ and $I_i \cap I_j = \emptyset$ for $i \neq j$ ordered pointwise.

From integrals for $t \in \mathcal{T}$ defined in the obvious way, e.g. as $\mathbf{E}(t) = \int_\Omega t(\omega) d\mu(\omega) = \sum_i \mu(I_i) c_i$, use lattice approximation(s):

$$\int_\Omega f(\omega) d\mu(\omega) = \bigsqcup \left\{ \int_\Omega t(\omega) d\mu(\omega) \mid t \in \mathcal{T} \ \wedge \ t \sqsubseteq f \right\}.$$

Integrals

For a measure space (Ω, σ, μ) one can define integral(s) for random variables, i.e. functions X or f, on Ω:

$$\mathbf{E}(f) = \int_\Omega f(\omega) d\mu(\omega)$$

Typically one starts with step functions $t \in \mathcal{T}$ with $t : \Omega \to \mathbb{R}$ which are constant on some base sets in $\mathcal{B}$ or the σ-algebra σ.

Example (Step functions on $[0, 1]$)

For $t = \sum_i t_i$ with $t_i(\omega) = c_i \in \mathbb{R}$ for ω in interval I_i s.t. $I_i = [a_i, b_i]$ and $\bigcup_i I_i = [0, 1]$ and $I_i \cap I_j = \emptyset$ for $i \neq j$ ordered pointwise.

From integrals for $t \in \mathcal{T}$ defined in the obvious way, e.g. as $\mathbf{E}(t) = \int_\Omega t(\omega) d\mu(\omega) = \sum_i \mu(I_i) c_i$, use lattice approximation(s):

$$\int_\Omega f(\omega) d\mu(\omega) = \bigsqcup \left\{ \int_\Omega t(\omega) d\mu(\omega) \mid t \in \mathcal{T} \land t \sqsubseteq f \right\}.$$

Integrals

For a measure space (Ω, σ, μ) one can define integral(s) for random variables, i.e. functions X or f, on Ω:

$$\mathbf{E}(f) = \int_{\Omega} f(\omega) d\mu(\omega)$$

Typically one starts with step functions $t \in \mathcal{T}$ with $t : \Omega \to \mathbb{R}$ which are constant on some base sets in $\mathcal{B}$ or the σ-algebra σ.

Example (Step functions on $[0, 1]$)

For $t = \sum_i t_i$ with $t_i(\omega) = c_i \in \mathbb{R}$ for ω in interval I_i s.t. $I_i = [a_i, b_i]$ and $\bigcup_i I_i = [0, 1]$ and $I_i \cap I_j = \emptyset$ for $i \neq j$ ordered pointwise.

From integrals for $t \in \mathcal{T}$ defined in the obvious way, e.g. as $\mathbf{E}(t) = \int_{\Omega} t(\omega) d\mu(\omega) = \sum_i \mu(I_i) c_i$, use lattice approximation(s):

$$\int_{\Omega} f(\omega) d\mu(\omega) = \bigsqcup \left\{ \int_{\Omega} t(\omega) d\mu(\omega) \mid t \in \mathcal{T} \ \wedge \ t \sqsubseteq f \right\}.$$

Integrals

For a measure space (Ω, σ, μ) one can define integral(s) for random variables, i.e. functions X or f, on Ω:

$$\mathbf{E}(f) = \int_\Omega f(\omega) d\mu(\omega)$$

Typically one starts with step functions $t \in \mathcal{T}$ with $t : \Omega \to \mathbb{R}$ which are constant on some base sets in $\mathcal{B}$ or the σ-algebra σ.

Example (Step functions on $[0, 1]$)

For $t = \sum_i t_i$ with $t_i(\omega) = c_i \in \mathbb{R}$ for ω in interval I_i s.t. $I_i = [a_i, b_i]$ and $\bigcup_i I_i = [0, 1]$ and $I_i \cap I_j = \emptyset$ for $i \neq j$ ordered pointwise.

From integrals for $t \in \mathcal{T}$ defined in the obvious way, e.g. as $\mathbf{E}(t) = \int_\Omega t(\omega) d\mu(\omega) = \sum_i \mu(I_i) c_i$, use lattice approximation(s):

$$\int_\Omega f(\omega) d\mu(\omega) = \bigsqcup \left\{ \int_\Omega t(\omega) d\mu(\omega) \mid t \in \mathcal{T} \ \wedge \ t \sqsubseteq f \right\}.$$

Integrals

For a measure space (Ω, σ, μ) one can define integral(s) for random variables, i.e. functions X or f, on Ω:

$$\mathbf{E}(f) = \int_\Omega f(\omega) d\mu(\omega)$$

Typically one starts with step functions $t \in \mathcal{T}$ with $t : \Omega \to \mathbb{R}$ which are constant on some base sets in $\mathcal{B}$ or the σ-algebra σ.

Example (Step functions on [0, 1])

For $t = \sum_i t_i$ with $t_i(\omega) = c_i \in \mathbb{R}$ for ω in interval I_i s.t. $I_i = [a_i, b_i]$ and $\bigcup_i I_i = [0, 1]$ and $I_i \cap I_j = \emptyset$ for $i \neq j$ ordered pointwise.

From integrals for $t \in \mathcal{T}$ defined in the obvious way, e.g. as $\mathbf{E}(t) = \int_\Omega t(\omega) d\mu(\omega) = \sum_i \mu(I_i) c_i$, use lattice approximation(s):

$$\int_\Omega f(\omega) d\mu(\omega) = \bigsqcup \left\{ \int_\Omega t(\omega) d\mu(\omega) \mid t \in \mathcal{T} \ \wedge \ t \sqsubseteq f \right\}.$$

Abstract Vector Spaces

Definition

A Vector Space (over a field $\mathbb{K}$, e.g. $\mathbb{R}$ or $\mathbb{C}$) is a set $\mathcal{V}$ together with two operations:

$$\text{Scalar Multiplication} \quad .\cdot. : \mathbb{K} \times \mathcal{V} \mapsto \mathcal{V}$$

$$\text{Vector Addition} \quad .+. : \mathcal{V} \times \mathcal{V} \mapsto \mathcal{V}$$

such that ($\forall x, y, z \in \mathcal{V}$ and $\alpha, \beta \in \mathbb{K}$):

1. $x + (y + z) = (x + y) + z$
2. $x + y = y + x$
3. $\exists o : x + o = x$
4. $\exists -x : x + (-x) = o$

1. $\alpha(x + y) = \alpha x + \alpha y$
2. $(\alpha + \beta)x = \alpha x + \beta x$
3. $(\alpha\beta)x = \alpha(\beta x)$
4. $1x = x$ $(1 \in \mathbb{K})$

Linear Operators

Definition

A map $\mathbf{T} : \mathcal{V} \to \mathcal{W}$ between two vector spaces $\mathcal{V}$ and $\mathcal{W}$ is called a linear map iff

1. $\mathbf{T}(x + y) = \mathbf{T}(x) + \mathbf{T}(y)$ and
2. $\mathbf{T}(\alpha x) = \alpha \mathbf{T}(x)$

for all $x, y \in \mathcal{V}$ and all $\alpha \in \mathbb{K}$ (e.g. $\mathbb{K} = \mathbb{C}$ or $\mathbb{R}$).

The set of all linear maps between $\mathcal{V}$ and $\mathcal{W}$ is denoted $\mathcal{L}(\mathcal{V}, \mathcal{W})$. For $\mathcal{V} = \mathcal{W}$ we talk about a linear operator on $\mathcal{V}$.

On normed vector spaces the continuous or equivalently bounded linear operators are of particular interest, i.e.

$$\mathcal{B}(\mathcal{V}) = \{\mathbf{T} \mid \|\mathbf{T}\| = \sup_{x \in \mathcal{V}} \frac{\|\mathbf{T}(x)\|}{\|x\|} < \infty\} \subseteq \mathcal{L}(\mathcal{V}) = \mathcal{L}(\mathcal{V}, \mathcal{V}).$$

Linear Operators

Definition

A map $\mathbf{T} : \mathcal{V} \to \mathcal{W}$ between two vector spaces $\mathcal{V}$ and $\mathcal{W}$ is called a linear map iff

1. $\mathbf{T}(x + y) = \mathbf{T}(x) + \mathbf{T}(y)$ and
2. $\mathbf{T}(\alpha x) = \alpha \mathbf{T}(x)$

for all $x, y \in \mathcal{V}$ and all $\alpha \in \mathbb{K}$ (e.g. $\mathbb{K} = \mathbb{C}$ or $\mathbb{R}$).

The set of all linear maps between $\mathcal{V}$ and $\mathcal{W}$ is denoted $\mathcal{L}(\mathcal{V}, \mathcal{W})$. For $\mathcal{V} = \mathcal{W}$ we talk about a linear operator on $\mathcal{V}$.

On normed vector spaces the continuous or equivalently bounded linear operators are of particular interest, i.e.

$$\mathcal{B}(\mathcal{V}) = \{\mathbf{T} \mid \|\mathbf{T}\| = \sup_{x \in \mathcal{V}} \frac{\|\mathbf{T}(x)\|}{\|x\|} < \infty\} \subseteq \mathcal{L}(\mathcal{V}) = \mathcal{L}(\mathcal{V}, \mathcal{V}).$$

Linear Operators

Definition

A map $\mathbf{T} : \mathcal{V} \to \mathcal{W}$ between two vector spaces $\mathcal{V}$ and $\mathcal{W}$ is called a linear map iff

1. $\mathbf{T}(x + y) = \mathbf{T}(x) + \mathbf{T}(y)$ and
2. $\mathbf{T}(\alpha x) = \alpha \mathbf{T}(x)$

for all $x, y \in \mathcal{V}$ and all $\alpha \in \mathbb{K}$ (e.g. $\mathbb{K} = \mathbb{C}$ or $\mathbb{R}$).

The set of all linear maps between $\mathcal{V}$ and $\mathcal{W}$ is denoted $\mathcal{L}(\mathcal{V}, \mathcal{W})$. For $\mathcal{V} = \mathcal{W}$ we talk about a linear operator on $\mathcal{V}$.

On normed vector spaces the continuous or equivalently bounded linear operators are of particular interest, i.e.

$$\mathcal{B}(\mathcal{V}) = \{\mathbf{T} \mid \|\mathbf{T}\| = \sup_{x \in \mathcal{V}} \frac{\|\mathbf{T}(x)\|}{\|x\|} < \infty\} \subseteq \mathcal{L}(\mathcal{V}) = \mathcal{L}(\mathcal{V}, \mathcal{V}).$$

Matrices and Lifted Functions

Any liner map $\mathbf{T} : \mathcal{V} \to \mathcal{W}$ can be conveniently be represented by a matrix, especially if they are finite dimensional; application then becomes vector/matrix multiplication.

Let $\{v_1, v_2, \ldots\}$ and $\{w_1, w_2, \ldots\}$ be bases for $\mathcal{V}$ and $\mathcal{W}$, then $\mathbf{T}$ can be represented via the matrix:

$$\mathbf{T} = (\mathbf{T}_{ij})_{ij} = \begin{pmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} & \cdots \\ \mathbf{T}_{21} & \mathbf{T}_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \text{ with } \mathbf{T}(v_i) = \sum_j \mathbf{T}_{ij} w_j.$$

Lifting Functions. Given a function $f : \Omega \to \Omega'$ then we can lift it to a linear map $\mathbf{T}_f : \mathcal{V}(\Omega) \to \mathcal{V}(\Omega')$:

$$\mathbf{T}_f(v_i) = f(v_i) \text{ i.e. } \mathbf{T}_{f,ij} = \begin{cases} 1 & \text{iff } f(v_i) = w_j \\ 0 & \text{otherwise} \end{cases}$$

Matrices and Lifted Functions

Any liner map $\mathbf{T} : \mathcal{V} \to \mathcal{W}$ can be conveniently be represented by a matrix, especially if they are finite dimensional; application then becomes vector/matrix multiplication.

Let $\{v_1, v_2, \ldots\}$ and $\{w_1, w_2, \ldots\}$ be bases for $\mathcal{V}$ and $\mathcal{W}$, then $\mathbf{T}$ can be represented via the matrix:

$$\mathbf{T} = (\mathbf{T}_{ij})_{ij} = \begin{pmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} & \cdots \\ \mathbf{T}_{21} & \mathbf{T}_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \text{ with } \mathbf{T}(v_i) = \sum_j \mathbf{T}_{ij} w_j.$$

Lifting Functions. Given a function $f : \Omega \to \Omega'$ then we can lift it to a linear map $\mathbf{T}_f : \mathcal{V}(\Omega) \to \mathcal{V}(\Omega')$:

$$\mathbf{T}_f(v_i) = f(v_i) \text{ i.e. } \mathbf{T}_{f,ij} = \begin{cases} 1 & \text{iff } f(v_i) = w_j \\ 0 & \text{otherwise} \end{cases}$$

Matrices and Lifted Functions

Any liner map $\mathbf{T} : \mathcal{V} \to \mathcal{W}$ can be conveniently be represented by a matrix, especially if they are finite dimensional; application then becomes vector/matrix multiplication.

Let $\{v_1, v_2, \ldots\}$ and $\{w_1, w_2, \ldots\}$ be bases for $\mathcal{V}$ and $\mathcal{W}$, then $\mathbf{T}$ can be represented via the matrix:

$$\mathbf{T} = (\mathbf{T}_{ij})_{ij} = \begin{pmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} & \cdots \\ \mathbf{T}_{21} & \mathbf{T}_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \text{ with } \mathbf{T}(v_i) = \sum_j \mathbf{T}_{ij} w_j.$$

Lifting Functions. Given a function $f : \Omega \to \Omega'$ then we can lift it to a linear map $\mathbf{T}_f : \mathcal{V}(\Omega) \to \mathcal{V}(\Omega')$:

$$\mathbf{T}_f(v_i) = f(v_i) \text{ i.e. } \mathbf{T}_{f,ij} = \begin{cases} 1 & \text{iff } f(v_i) = w_j \\ 0 & \text{otherwise} \end{cases}$$

Metric Spaces

Vector spaces are purely algebraic structures. One can also equip them with a topological structure. For finite dimensional vector spaces the topology is essentially unique, for infinite dimensional spaces one often defines a metric topology.

Definition

A metric space is a set Ω and a real-valued function $d(.,.)$, a metric, on $\Omega \times \Omega$ which satisfies:

1. $d(x, y) \geq 0$
2. $d(x, y) = 0 \iff x = y$
3. $d(x, y) = d(y, x)$
4. $d(x, z) \leq d(x, y) + d(y, z)$

Metric Spaces

Vector spaces are purely algebraic structures. One can also equip them with a topological structure. For finite dimensional vector spaces the topology is essentially unique, for infinite dimensional spaces one often defines a metric topology.

Definition

A metric space is a set Ω and a real-valued function $d(.,.)$, a metric, on $\Omega \times \Omega$ which satisfies:

1. $d(x, y) \geq 0$
2. $d(x, y) = 0 \iff x = y$
3. $d(x, y) = d(y, x)$
4. $d(x, z) \leq d(x, y) + d(y, z)$

Complete Metric Spaces

In a metric space we can define a basis for the topology open sets via open balls, i.e. sets $B(x, \varepsilon) = \{x' \mid d(x, x') < \varepsilon\}$, i.e. open sets are those which are unions of open balls.

Given a sequence $(x_i)_{i \in \mathbb{N}}$ of points in a topological space. We say that it converges if there exists $x = \lim x_i$ such that for all neighbourhoods $U(x)$ of x there $\exists N$ s.t. for $n > N : x_n \in U(x)$.

A sequence of elements $(x_i)_{i \in \mathbb{N}}$ in a metric space (X, d) is called a Cauchy sequence if

$$\forall \varepsilon > 0 \; \exists N : n, m \geq N \Rightarrow d(x_n, x_m) < \varepsilon.$$

A metric space (X, d) in which all Cauchy sequences converge is called complete (metric) space.

Complete Metric Spaces

In a metric space we can define a basis for the topology open sets via open balls, i.e. sets $B(x, \varepsilon) = \{x' \mid d(x, x') < \varepsilon\}$, i.e. open sets are those which are unions of open balls.

Given a sequence $(x_i)_{i \in \mathbb{N}}$ of points in a topological space. We say that it converges if there exists $x = \lim x_i$ such that for all neighbourhoods $U(x)$ of x there $\exists N$ s.t. for $n > N : x_n \in U(x)$.

A sequence of elements $(x_i)_{i \in \mathbb{N}}$ in a metric space (X, d) is called a Cauchy sequence if

$$\forall \varepsilon > 0 \ \exists N : n, m \geq N \Rightarrow d(x_n, x_m) < \varepsilon.$$

A metric space (X, d) in which all Cauchy sequences converge is called complete (metric) space.

Complete Metric Spaces

In a metric space we can define a basis for the topology open sets via open balls, i.e. sets $B(x, \varepsilon) = \{x' \mid d(x, x') < \varepsilon\}$, i.e. open sets are those which are unions of open balls.

Given a sequence $(x_i)_{i \in \mathbb{N}}$ of points in a topological space. We say that it converges if there exists $x = \lim x_i$ such that for all neighbourhoods $U(x)$ of x there $\exists N$ s.t. for $n > N : x_n \in U(x)$.

A sequence of elements $(x_i)_{i \in \mathbb{N}}$ in a metric space (X, d) is called a Cauchy sequence if

$$\forall \varepsilon > 0 \; \exists N : n, m \geq N \Rightarrow d(x_n, x_m) < \varepsilon.$$

A metric space (X, d) in which all Cauchy sequences converge is called complete (metric) space.

Complete Metric Spaces

In a metric space we can define a basis for the topology open sets via open balls, i.e. sets $B(x, \varepsilon) = \{x' \mid d(x, x') < \varepsilon\}$, i.e. open sets are those which are unions of open balls.

Given a sequence $(x_i)_{i \in \mathbb{N}}$ of points in a topological space. We say that it converges if there exists $x = \lim x_i$ such that for all neighbourhoods $U(x)$ of x there $\exists N$ s.t. for $n > N : x_n \in U(x)$.

A sequence of elements $(x_i)_{i \in \mathbb{N}}$ in a metric space (X, d) is called a Cauchy sequence if

$$\forall \varepsilon > 0 \; \exists N : n, m \geq N \Rightarrow d(x_n, x_m) < \varepsilon.$$

A metric space (X, d) in which all Cauchy sequences converge is called complete (metric) space.

Banach Spaces

Definition

A complex vector space $\mathcal{V}$ is called a normed (vector) space if there is a real valued function $\|.\|$ on $\mathcal{V}$ that satisfies ($\forall x, y \in \mathcal{V}$ and $\forall \alpha \in \mathbb{C}$):

1. $\|x\| \geq 0$
2. $\|x\| = 0 \iff x = o$
3. $\|\alpha x\| = |\alpha| \, \|x\|$
4. $\|x + y\| \leq \|x\| + \|y\|$

The function $\|.\|$ is called a norm on $\mathcal{V}$.

We have a Banach space if the topology induced by $d(x, y) = \|x - y\|$ is complete – always for finite dimensional spaces.

Banach Spaces

Definition

A complex vector space $\mathcal{V}$ is called a normed (vector) space if there is a real valued function $\|.\|$ on $\mathcal{V}$ that satisfies ($\forall x, y \in \mathcal{V}$ and $\forall \alpha \in \mathbb{C}$):

1. $\|x\| \geq 0$
2. $\|x\| = 0 \iff x = o$
3. $\|\alpha x\| = |\alpha| \, \|x\|$
4. $\|x + y\| \leq \|x\| + \|y\|$

The function $\|.\|$ is called a norm on $\mathcal{V}$.

We have a Banach space if the topology induced by $d(x, y) = \|x - y\|$ is complete – always for finite dimensional spaces.

Hilbert Spaces

Definition

A complex vector space $\mathcal{H}$ is called an inner product space (or (pre-)Hilbert space) if there is a complex valued function $\langle ., . \rangle$ on $\mathcal{H} \times \mathcal{H}$ that satisfies ($\forall x, y, z \in \mathcal{H}$ and $\forall \alpha \in \mathbb{C}$):

1. $\langle x, x \rangle \geq 0$
2. $\langle x, x \rangle = 0 \iff x = o$
3. $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$
4. $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$
5. $\langle x, y \rangle = \overline{\langle y, x \rangle}$

The function $\langle ., . \rangle$ is called an inner product on $\mathcal{H}$.

If the topology induced by $\|x\| = \sqrt{\langle x, x \rangle}$ is complete then we have a Hilbert space – always for finite dimensional spaces.

Dual Spaces

Linear functionals on a vector space $\mathcal{V}$ are maps $f : \mathcal{V} \to \mathbb{K}$ with $f(x + y) = f(x) + f(y)$ and $f(\alpha x) = \alpha f(x)$ for all $x, y \in \mathcal{V}, \alpha \in \mathbb{K}$.

Theorem (Riesz Representation Theorem)

Every (bounded) linear functional on a Hilbert space $\mathcal{H}$ can be represented by a vector in the Hilbert space $\mathcal{H}$, such that

$$f(x) = \langle y_f | x \rangle = f_y(x)$$

The dual Hilbert space $\mathcal{H}^*$ is isomorphic to the original Hilbert space $\mathcal{H}$, e.g. for the universal Hilbert space $\ell_2(\mathbb{N})^* = \ell_2(\mathbb{N})$.

$$\ell_p(\Omega) = \left\{ (x_i)_{i\in\Omega} \mid \left(\sum_{i\in\Omega} |x_i|^p \right)^{\frac{1}{p}} < \infty \right\}$$

Dual Spaces

Linear functionals on a vector space $\mathcal{V}$ are maps $f : \mathcal{V} \to \mathbb{K}$ with $f(x + y) = f(x) + f(y)$ and $f(\alpha x) = \alpha f(x)$ for all $x, y \in \mathcal{V}, \alpha \in \mathbb{K}$.

Theorem (Riesz Representation Theorem)

Every (bounded) linear functional on a Hilbert space $\mathcal{H}$ can be represented by a vector in the Hilbert space $\mathcal{H}$, such that

$$f(x) = \langle y_f | x \rangle = f_y(x)$$

The dual Hilbert space $\mathcal{H}^*$ is isomorphic to the original Hilbert space $\mathcal{H}$, e.g. for the universal Hilbert space $\ell_2(\mathbb{N})^* = \ell_2(\mathbb{N})$.

$$\ell_p(\Omega) = \left\{ (x_i)_{i \in \Omega} \mid \left(\sum_{i \in \Omega} |x_i|^p \right)^{\frac{1}{p}} < \infty \right\}$$

Dual Spaces

Linear functionals on a vector space $\mathcal{V}$ are maps $f : \mathcal{V} \to \mathbb{K}$ with $f(x+y) = f(x) + f(y)$ and $f(\alpha x) = \alpha f(x)$ for all $x, y \in \mathcal{V}, \alpha \in \mathbb{K}$.

Theorem (Riesz Representation Theorem)

Every (bounded) linear functional on a Hilbert space $\mathcal{H}$ can be represented by a vector in the Hilbert space $\mathcal{H}$, such that

$$f(x) = \langle y_f | x \rangle = f_y(x)$$

The dual Hilbert space $\mathcal{H}^*$ is isomorphic to the original Hilbert space $\mathcal{H}$, e.g. for the universal Hilbert space $\ell_2(\mathbb{N})^* = \ell_2(\mathbb{N})$.

$$\ell_p(\Omega) = \left\{ (x_i)_{i \in \Omega} \mid \left(\sum_{i \in \Omega} |x_i|^p \right)^{\frac{1}{p}} < \infty \right\}$$

Dual Spaces

Linear functionals on a vector space $\mathcal{V}$ are maps $f : \mathcal{V} \to \mathbb{K}$ with $f(x+y) = f(x) + f(y)$ and $f(\alpha x) = \alpha f(x)$ for all $x, y \in \mathcal{V}, \alpha \in \mathbb{K}$.

Theorem (Riesz Representation Theorem)

Every (bounded) linear functional on a Hilbert space $\mathcal{H}$ can be represented by a vector in the Hilbert space $\mathcal{H}$, such that

$$f(x) = \langle y_f | x \rangle = f_y(x)$$

The dual Hilbert space $\mathcal{H}^*$ is isomorphic to the original Hilbert space $\mathcal{H}$, e.g. for the universal Hilbert space $\ell_2(\mathbb{N})^* = \ell_2(\mathbb{N})$.

$$L_p(\Omega) = \left\{ f : \Omega \to \mathbb{K} \mid \left(\int_\Omega |f(\omega)|^p d\mu(\omega) \right)^{\frac{1}{p}} < \infty \right\}$$

Measure Theory and Duality

Classical Banach and Hilbert spaces are the sequence spaces $\ell_1(\mathbb{N}), \ell_2(\mathbb{N}), \ldots, \ell_\infty(\mathbb{N})$ or the spaces $L_1(\Omega), L_2(\Omega), \ldots, L_\infty(\Omega)$ of (equivalence classes) of integrable function $f : \Omega \to \mathbb{R}$ for (general) Ω. Then ℓ_p/ℓ_q or L_p/L_q are dual for $\frac{1}{p} + \frac{1}{q} = 1$.

There is a general duality between vectors and functionals. This duality corresponds to the duality between random variables/functions and distributions/measures. One can identify expectation values, integrals and inner products:

$$\mathbf{E}(f, \mu) = \int_\Omega f(\omega) d\mu(\omega) = \langle f|\mu \rangle .$$

Example

Consider the set $L_\infty(\Omega)$ of bounded functions on Ω with $\|f\|_\infty = \sup_\Omega |f(\omega)|$ and the dual space $L_1(\Omega)$ of "measures".

Measure Theory and Duality

Classical Banach and Hilbert spaces are the sequence spaces $\ell_1(\mathbb{N}), \ell_2(\mathbb{N}), \ldots, \ell_\infty(\mathbb{N})$ or the spaces $L_1(\Omega), L_2(\Omega), \ldots, L_\infty(\Omega)$ of (equivalence classes) of integrable function $f : \Omega \to \mathbb{R}$ for (general) Ω. Then ℓ_p/ℓ_q or L_p/L_q are dual for $\frac{1}{p} + \frac{1}{q} = 1$.

There is a general duality between vectors and functionals. This duality corresponds to the duality between random variables/functions and distributions/measures. One can identify expectation values, integrals and inner products:

$$\mathbf{E}(f, \mu) = \int_\Omega f(\omega) d\mu(\omega) = \langle f | \mu \rangle .$$

Example

Consider the set $L_\infty(\Omega)$ of bounded functions on Ω with $\|f\|_\infty = \sup_\Omega |f(\omega)|$ and the dual space $L_1(\Omega)$ of "measures".

Measure Theory and Duality

Classical Banach and Hilbert spaces are the sequence spaces $\ell_1(\mathbb{N}), \ell_2(\mathbb{N}), \ldots, \ell_\infty(\mathbb{N})$ or the spaces $L_1(\Omega), L_2(\Omega), \ldots, L_\infty(\Omega)$ of (equivalence classes) of integrable function $f : \Omega \to \mathbb{R}$ for (general) Ω. Then ℓ_p/ℓ_q or L_p/L_q are dual for $\frac{1}{p} + \frac{1}{q} = 1$.

There is a general duality between vectors and functionals. This duality corresponds to the duality between random variables/functions and distributions/measures. One can identify expectation values, integrals and inner products:

$$\mathbf{E}(f, \mu) = \int_\Omega f(\omega) d\mu(\omega) = \langle f | \mu \rangle .$$

Example

Consider the set $L_\infty(\Omega)$ of bounded functions on Ω with $\|f\|_\infty = \sup_\Omega |f(\omega)|$ and the dual space $L_1(\Omega)$ of "measures".

Measure Theory and Duality

Classical Banach and Hilbert spaces are the sequence spaces $\ell_1(\mathbb{N}), \ell_2(\mathbb{N}), \ldots, \ell_\infty(\mathbb{N})$ or the spaces $L_1(\Omega), L_2(\Omega), \ldots, L_\infty(\Omega)$ of (equivalence classes) of integrable function $f : \Omega \to \mathbb{R}$ for (general) Ω. Then ℓ_p/ℓ_q or L_p/L_q are dual for $\frac{1}{p} + \frac{1}{q} = 1$.

There is a general duality between vectors and functionals. This duality corresponds to the duality between random variables/functions and distributions/measures. One can identify expectation values, integrals and inner products:

$$\mathbf{E}(f, \mu) = \int_\Omega f(\omega) d\mu(\omega) = \langle f | \mu \rangle .$$

Example

Consider the set $L_\infty(\Omega)$ of bounded functions on Ω with $\|f\|_\infty = \sup_\Omega |f(\omega)|$ and the dual space $L_1(\Omega)$ of "measures".

Measure Theory and Duality

Classical Banach and Hilbert spaces are the sequence spaces $\ell_1(\mathbb{N}), \ell_2(\mathbb{N}), \ldots, \ell_\infty(\mathbb{N})$ or the spaces $L_1(\Omega), L_2(\Omega), \ldots, L_\infty(\Omega)$ of (equivalence classes) of integrable function $f : \Omega \to \mathbb{R}$ for (general) Ω. Then ℓ_p/ℓ_q or L_p/L_q are dual for $\frac{1}{p} + \frac{1}{q} = 1$.

There is a general duality between vectors and functionals. This duality corresponds to the duality between random variables/functions and distributions/measures. One can identify expectation values, integrals and inner products:

$$\mathbf{E}(f, \mu) = \int_\Omega f(\omega) d\mu(\omega) = \langle f|\mu \rangle .$$

Example

Consider the set $L_\infty(\Omega)$ of bounded functions on Ω with $\|f\|_\infty = \sup_\Omega |f(\omega)|$ and the dual space $L_1(\Omega)$ of "measures".

Measure Theory and Duality

Classical Banach and Hilbert spaces are the sequence spaces $\ell_1(\mathbb{N}), \ell_2(\mathbb{N}), \ldots, \ell_\infty(\mathbb{N})$ or the spaces $L_1(\Omega), L_2(\Omega), \ldots, L_\infty(\Omega)$ of (equivalence classes) of integrable function $f : \Omega \to \mathbb{R}$ for (general) Ω. Then ℓ_p/ℓ_q or L_p/L_q are dual for $\frac{1}{p} + \frac{1}{q} = 1$.

There is a general duality between vectors and functionals. This duality corresponds to the duality between random variables/functions and distributions/measures. One can identify expectation values, integrals and inner products:

$$\mathbf{E}(f, \mu) = \int_\Omega f(\omega) d\mu(\omega) = \langle f | \mu \rangle .$$

Example

Consider the set $L_\infty(\Omega)$ of bounded functions on Ω with $\|f\|_\infty = \sup_\Omega |f(\omega)|$ and the dual space $L_1(\Omega)$ of "measures".

Some Text Books

- David Applebaum: *Probability and Information*, Cambridge University Press, 1996/2008.
- David Strizacker: *Probability and Random Variables*, Cambridge University Press, 1996/2008.
- Patrick Billingsley: *Probability and Measure* John Wiley & Sons, 1979.
- Carlos Kubrusly: *The Elements of Operator Theory*, Birkhäuser, 2011.
- Achim Klenke: *Probability Theory - A Comprehensive Course*, Springer Verlag, 2006.
- Steven Roman: *Advanced Linear Algebra*, Springer Verlag, 2005.
- Kadison and Ringrose: *Fundamentals of the Theory of Operator Algebras*, AMS, 1997.

Some Text Books

- David Applebaum: *Probability and Information*, Cambridge University Press, 1996/2008.
- David Strizacker: *Probability and Random Variables*, Cambridge University Press, 1996/2008.
- Patrick Billingsley: *Probability and Measure* John Wiley & Sons, 1979.
- Carlos Kubrusly: *The Elements of Operator Theory*, Birkhäuser, 2011.
- Achim Klenke: *Probability Theory - A Comprehensive Course*, Springer Verlag, 2006.
- Steven Roman: *Advanced Linear Algebra*, Springer Verlag, 2005.
- Kadison and Ringrose: *Fundamentals of the Theory of Operator Algebras*, AMS, 1997.

Some Text Books

- David Applebaum: *Probability and Information*, Cambridge University Press, 1996/2008.
- David Strizacker: *Probability and Random Variables*, Cambridge University Press, 1996/2008.
- Patrick Billingsley: *Probability and Measure* John Wiley & Sons, 1979.
- Carlos Kubrusly: *The Elements of Operator Theory*, Birkhäuser, 2011.
- Achim Klenke: *Probability Theory - A Comprehensive Course*, Springer Verlag, 2006.
- Steven Roman: *Advanced Linear Algebra*, Springer Verlag, 2005.
- Kadison and Ringrose: *Fundamentals of the Theory of Operator Algebras*, AMS, 1997.

Some Text Books

- David Applebaum: *Probability and Information*, Cambridge University Press, 1996/2008.
- David Strizacker: *Probability and Random Variables*, Cambridge University Press, 1996/2008.
- Patrick Billingsley: *Probability and Measure* John Wiley & Sons, 1979.
- Carlos Kubrusly: *The Elements of Operator Theory*, Birkhäuser, 2011.
- Achim Klenke: *Probability Theory - A Comprehensive Course*, Springer Verlag, 2006.
- Steven Roman: *Advanced Linear Algebra*, Springer Verlag, 2005.
- Kadison and Ringrose: *Fundamentals of the Theory of Operator Algebras*, AMS, 1997.

Some Text Books

- David Applebaum: *Probability and Information*, Cambridge University Press, 1996/2008.
- David Strizacker: *Probability and Random Variables*, Cambridge University Press, 1996/2008.
- Patrick Billingsley: *Probability and Measure* John Wiley & Sons, 1979.
- Carlos Kubrusly: *The Elements of Operator Theory*, Birkhäuser, 2011.
- Achim Klenke: *Probability Theory - A Comprehensive Course*, Springer Verlag, 2006.
- Steven Roman: *Advanced Linear Algebra*, Springer Verlag, 2005.
- Kadison and Ringrose: *Fundamentals of the Theory of Operator Algebras*, AMS, 1997.

Some Text Books

- David Applebaum: *Probability and Information*, Cambridge University Press, 1996/2008.
- David Strizacker: *Probability and Random Variables*, Cambridge University Press, 1996/2008.
- Patrick Billingsley: *Probability and Measure* John Wiley & Sons, 1979.
- Carlos Kubrusly: *The Elements of Operator Theory*, Birkhäuser, 2011.
- Achim Klenke: *Probability Theory - A Comprehensive Course*, Springer Verlag, 2006.
- Steven Roman: *Advanced Linear Algebra*, Springer Verlag, 2005.
- Kadison and Ringrose: *Fundamentals of the Theory of Operator Algebras*, AMS, 1997.

Some Text Books

- David Applebaum: *Probability and Information*, Cambridge University Press, 1996/2008.
- David Strizacker: *Probability and Random Variables*, Cambridge University Press, 1996/2008.
- Patrick Billingsley: *Probability and Measure* John Wiley & Sons, 1979.
- Carlos Kubrusly: *The Elements of Operator Theory*, Birkhäuser, 2011.
- Achim Klenke: *Probability Theory - A Comprehensive Course*, Springer Verlag, 2006.
- Steven Roman: *Advanced Linear Algebra*, Springer Verlag, 2005.
- Kadison and Ringrose: *Fundamentals of the Theory of Operator Algebras*, AMS, 1997.