Reviews of Probability Theory and Convex Analysis

Operations Research

Anthony Papavasiliou

Short Reviews

Table of Contents

Sigma Algebra

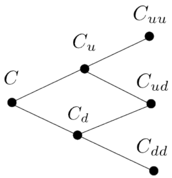
Given a **sample space** Ω, a sigma algebra **sigma-algebra** $\mathcal{A}$ is a set of subsets of Ω such that

- $\Omega \in \mathcal{A}$
- if $A \in \mathcal{A}$ then also $\Omega - A \in \mathcal{A}$
- if $A_i \in \mathcal{A}$ for $i = 1, 2, \ldots$ then also $\cup_{i=1}^{\infty} A_i \in \mathcal{A}$

Sigma Algebras for Markov Decision Processes

- Given a sample space Ω, there is no unique sigma-algebra of Ω, here are two
  - $\{\emptyset, \Omega\}$
  - 2^{Ω} set of all subsets (**power set**) of Ω
- In these notes we will focus on finite Ω, and its power set, denoted $\mathcal{B}(\Omega)$
- Elements of a sigma-algebra are called **events**
- For a Markov decision process with finite time periods, $t = 0, \ldots, H$, denote set of states as S_t, set of actions as A_t
- Gradual revelation of information is modeled by defining, for each stage t, a sigma-algebra which is a subset of the power set of $\Omega = (S_0, A_0) \times \ldots \times (S_{H-1}, A_{H-1}) \times S_H$

Example: Stock Price Evolution



- State space is the set of values that the stock price can take at each stage: $S_0 = \{C\}$, $S_1 = \{C_u, C_d\}$, $S_2 = \{C_{uu}, C_{ud}, C_{dd}\}$
- Sample space is

$$\Omega = S_0 \times S_1 \times S_2 = \{(C, C_u, C_{uu}), (C, C_u, C_{ud}), (C, C_u, C_{dd}), (C, C_d, C_{ud}), (C, C_d, C_{dd})\}$$

Information of Stock Price Evolution in Period 2

Information in period 2:

$$\Omega_2 = \{(C, C_u, C_{uu}), (C, C_u, C_{ud}), (C, C_d, C_{ud}), (C, C_d, C_{dd})\}$$

$$\mathcal{B}(\Omega_2) =$$

$$\{\emptyset, \{(C, C_u, C_{uu})\}, \{(C, C_u, C_{ud})\}, \{(C, C_d, C_{du})\}, \{(C, C_d, C_{dd})\}$$

$$\{(C, C_u, C_{uu}), (C, C_u, C_{ud})\}, \{(C, C_u, C_{uu}), (C, C_d, C_{ud})\},$$

$$\{(C, C_u, C_{uu}), (C, C_d, C_{dd})\}, \{(C, C_u, C_{ud}), (C, C_d, C_{ud})\},$$

$$\{(C, C_u, C_{ud}), (C, C_d, C_{dd})\}, \{(C, C_d, C_{ud}), (C, C_d, C_{dd})\},$$

$$\{(C, C_u, C_{uu}), (C, C_u, C_{ud}), (C, C_d, C_{ud})\},$$

$$\{(C, C_u, C_{uu}), (C, C_u, C_{ud}), (C, C_d, C_{dd})\},$$

$$\{(C, C_u, C_{uu}), (C, C_d, C_{ud}), (C, C_d, C_{dd})\}$$

$$\{(C, C_u, C_{ud}), (C, C_d, C_{ud}), (C, C_d, C_{dd})\}\}.$$

$$\{(C, C_u, C_{uu}), (C, C_u, C_{ud}), (C, C_d, C_{ud}), (C, C_d, C_{dd})\},$$

Information of Stock Price Evolution in Period 2

'the stock price in period 2 is C_{ud}': identifiable (corresponds to $\{(C, C_u, C_{ud}), (C, C_d, C_{ud})\}$, which is an element of $\mathcal{B}(\Omega)$)

Information of Stock Price Evolution in Period 0

Information in period 0:

$$\mathcal{A}_0 = \{\emptyset, \Omega\}.$$

This is a valid sigma-algebra on Ω (satisfies all three conditions of the definition of a sigma-algebra)

Information of Stock Price Evolution in Period 1

Information in period 1:

$$\begin{aligned}
&\mathcal{A}_1 = \{\emptyset, \\
&\{(C, C_u, C_{uu}), (C, C_u, C_{ud}), (C, C_u, C_{dd})\}, \\
&\{(C, C_d, C_{uu}), (C, C_d, C_{ud}), (C, C_d, C_{dd})\}, \\
&\Omega\}
\end{aligned}$$

- 'the stock price in period 0 is C, and in period 1 it is C_u': distinguishable (2nd element in $\mathcal{A}_1$)
- 'the stock price in period 0 was C, in period 1 it is C_u, and in period 2 it is C_{uu}': not distinguishable (not in $\mathcal{A}_2$)

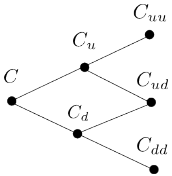
Probability Space

A **measurable probability space** is the triplet $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$, where Ω is the sample space, $\mathcal{B}(\Omega)$ is the set of all subsets of Ω, and $\mathbb{P} : \mathcal{B}(\Omega) \to [0, 1]$ is the probability measure that obeys the following properties:

- $\mathbb{P}(\emptyset) = 0$,
- $\mathbb{P}(\Omega) = 1$, and
- $\mathbb{P}(A_1 \cup A_2) = P(A_1) + P(A_2)$ if $A_1 \cap A_2 = \emptyset$

Note: $\mathbb{P}[\cdot]$ and $\mathbb{P}(\cdot)$ will be used interchangeably

Example: Stock Price Evolution



Suppose that probability of upward movement in each time stage is equal to 0.6

Probability measure of period 0: $\mathbb{P}_0[(\emptyset] = 0$ and $\mathbb{P}_0[\Omega] = 1$

Probability measure for stage 1:

$$\mathbb{P}_1[\emptyset] = 0$$

$$\mathbb{P}_1[\{(C, C_u, C_{uu}), (C, C_u, C_{ud}), (C, C_u, C_{dd})\}] = 0.6$$

$$\mathbb{P}_1[\{(C, C_d, C_{uu}), (C, C_d, C_{ud}), (C, C_d, C_{dd})\}] = 0.4$$

$$\mathbb{P}_1[\Omega] = 1$$

Filtration

Given $(\Omega, \mathcal{B}(\Omega))$, a **filtration** is an increasing sequence of sigma-algebras $\{\mathcal{A}_t\}_{t \geq 0}$ where each t is non-negative and

$$t_1 \leq t_2 \Rightarrow \mathcal{A}_{t_1} \subseteq A_{t_2}$$

In the stock price example, the sequence $(\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2)$, where $\mathcal{A}_2 = \mathcal{B}(\Omega)$, defines a filtration on $(\Omega, \mathcal{B}(\Omega))$

Specifying a Probability Measure in Filtrations

Any event $A_1 \in \mathcal{A}_{t_1}$ can be described as the union of disjoint events in $\mathcal{A}_{t_2}$, i.e. $A_1 = B_1 \cup \ldots \cup B_n$ where n is finite, $B_i \cap B_j = \emptyset$ for $i \neq j$, and $B_i \in \mathcal{A}_{t_2}, i = 1, \ldots, n$

Conclusion: the definition of a probability measure $\mathbb{P}_{t_2}$ on $\mathcal{A}_{t_2}$ implies a probability measure for all $\mathcal{A}_{t_1}$ with $t_1 \leq t_2$

In particular, $\mathbb{P}_{t_1}[A_1] = \mathbb{P}_{t_1}[B_1 \cup \ldots B_n] = \sum_{i=1}^{n} \mathbb{P}_{t_2}[B_i]$ for any $A \in \mathcal{A}_1$

Example: Stock Price Evolution

The probability measure on $(\Omega, \mathcal{A}_1)$ which is derived previously implies a probability measure $\mathbb{P}_0$ on $\mathcal{A}_0$. To see this, note that

$$
\begin{aligned}
&\mathbb{P}_0[\Omega] = \\
&\mathbb{P}_1[\{(C, C_u, C_{uu}), (C, C_u, C_{ud}), (C, C_u, C_{dd})\}] + \\
&\mathbb{P}_1[\{(C, C_d, C_{uu}), (C, C_d, C_{ud}), (C, C_d, C_{dd})\}]
\end{aligned}
$$

Conditional Probability

The **conditional probability** of event A given event B is defined as

$$\mathbb{P}[A|B] = \begin{cases} \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}, & \mathbb{P}[B] > 0 \\ 0, & \mathbb{P}[B] = 0 \end{cases}$$

Stock Pricing Example

Random variables in stock pricing example: price ξ_t in stage t

For period 0,

$$\xi_0(\omega) = C, \omega \in \Omega$$

For period 1,

$$\xi_1(\omega) = C_u, \omega = (C, C_u, \cdot)$$
$$\xi_1(\omega) = C_d, \omega = (C, C_d, \cdot)$$

For period 2,

$$\xi_2(\omega) = C_{uu}, \omega = (C, \cdot, C_{uu})$$
$$\xi_2(\omega) = C_{ud}, \omega = (C, \cdot, C_{ud})$$
$$\xi_2(\omega) = C_{ud}, \omega = (C, \cdot, C_{ud})$$

Measurable Function

Informal definition: a function is **measurable** with respect to a sigma-algebra, if it can be defined on the basis of the information that is available in the sigma-algebra

Stock pricing example

- ξ_2 is *not* measurable with respect to $\mathcal{A}_1$, because by observing a specific value of ξ_2 one can distinguish between, for example, (C, C_u, C_{uu}) and (C, C_u, C_{ud})
- In mathematical terms, $\{\omega \in \Omega : \xi_2(\omega) = C_{uu}\} \notin \mathcal{A}_1$, which means that the amount of information needed to define ξ_2 is not available in $\mathcal{A}_1$, therefore ξ_2 is not measurable with respect to $\mathcal{A}_1$

Random Variable, Random Vector, Stochastic Process

A **random variable** $\xi : \Omega \to \mathbb{R}$ is a function that maps random outcomes to real values

A **random vector** is a function $\xi : \Omega \to \mathbb{R}^n$ that maps outcomes to real-valued vectors

Given an index set T, and a probability space $(\Omega, \mathcal{B}(A), \mathbb{P})$, a **stochastic process** is a collection of $\mathbb{R}^n$-valued random vectors, which can be written as $(X(t) : t \in T)$

Distribution Functions

The **cumulative distribution function** of a random variable ξ is defined as $F_\xi(x) = P(\xi \le x)$

For discrete random variables, the **probability mass function** f is defined as $f(\xi^k) = P(\xi = \xi^k), k \in K$ with $\sum_{k \in K} f(\xi^k) = 1$

For continuous random variables, the **density function** f is defined by $P(a \le \xi \le b) = \int_a^b f(\xi) d\xi = \int_a^b dF(\xi)$ with $\int_{-\infty}^{\infty} dF(\xi) = 1$

Common Statistics of Random Variables

The **expectation** of a random variable is defined as $\mu = \sum_{k \in K} \xi^k f(\xi^k)$ for discrete random variables and as $\int_{-\infty}^{\infty} \xi dF(\xi)$ continuous random variables
The moment **rth moment** of ξ is $\bar{\xi}^{(r)} = \mathbb{E}[\xi^r]$

The **variance** of a random variable is defined as $\mathbb{E}[(\xi - \mu)^2]$

The α**-quantile** of ξ is a point η such that for $0 < \alpha < 1$, $\eta = \min\{x | F(x) \geq \alpha\}$

Convergence in Distribution

A sequence $X_1, X_2, \ldots$ of random variables is said to **converge in distribution**, or **converge weakly**, or **converge in law** to a random variable X if

$$\lim_{n\to\infty} F_n(x) = F(x)$$

for every $x \in \mathbb{R}$ at which F is continuous. F and F_n are the cumulative distribution functions of X and X_n respectively

Table of Contents

Polyhedron

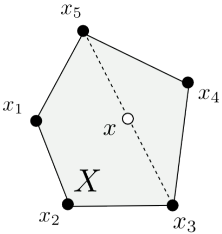
A **polyhedron** is a set in $\mathbb{R}^n$ which can be expressed as $\{x : Ax \leq b\}$, where $A \in \mathbb{R}^m \times \mathbb{R}^n$ and $b \in \mathbb{R}^m$.

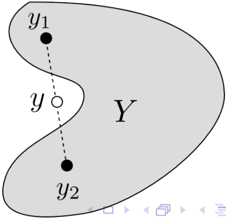
Convex

Consider a set of points $x_i \in \mathbb{R}^n, i = 1, \ldots, n$, a **convex combination** of these points is a point $\sum_{i=1}^n \lambda_i x_i$, such that $\sum_{i=1}^n \lambda_i = 1$ and $\lambda_i \geq 0, i = 1, \ldots, n$

X is a **convex set** if it contains any convex combination of points $x_i \in X$

The **convex hull** of a set of points is the set of all convex combinations of these points





Extreme Points, Extreme Rays

An **extreme point** of a convex set is a point which cannot be expressed as the convex combination of two distinct points in the set

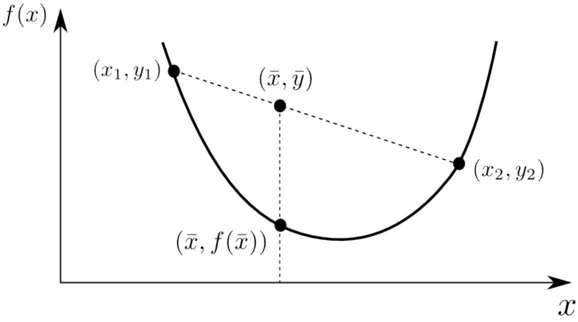
A point $r \in \mathbb{R}^n$ is a **ray** of a polyhedron P if and only if for any point $x \in P$, $\{y \in \mathbb{R}^n : y = x + \lambda r, \lambda \geq 0\} \subseteq P$

A ray r of P is an **extreme ray** if it cannot be expressed as a convex combination of other rays of P

Convex and Concave Functions

f is a **convex function** if for all $0 \leq \lambda \leq 1$ and any x_1, x_2 we have $f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2)$

f is **concave** if $-f$ is convex.



Frequently Encountered Classes of Functions

f is an **additively separable function** if it can be written as $f(x_1, x_2, \ldots, x_n) = f_1(x_1) + f_2(x_2) + \ldots + f_n(x_n)$

The **domain** of f, dom f, is the set where f is finite

A continuous function f is **piecewise linear** if it can be written as

$$f(x) = \max_{i=1,\ldots,n} (a_i^T x + b_i)$$

for all $x \in \text{dom } f$, where $a_i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$, and n a finite integer number

Convex Optimization Problems

An **optimization problem** is the problem of finding the minimum of a function f over a set $X \subset \mathbb{R}^n$:

$$\min f(x)$$
$$\text{subject to } x \in X$$

X is the **feasible set** of the problem, f is the **objective function** of the problem

Any $x \in X$ is a **feasible solution**, any $x^\star \in X$ such that $f(x^\star) \leq f(x)$ for any $x \in X$ is an **optimal solution**

A **convex optimization problem** is an optimization problem with a convex objective function and a convex set of feasible solutions

Supporting Hyperplane

A supporting hyperplane **supporting hyperplane** of a function $f(x)$ at x_0 is a set of points $a^T x + b$ such that $f(x_0) = a^T x_0 + b$ and $f(x) \geq a^T x + b$ for all x