CSE 840: Computational Foundations of Artificial Intelligence March 12, 2025

σ-Algebra

Instructor: Vishnu Boddeti *Scribe: Aafnan Mahmood*

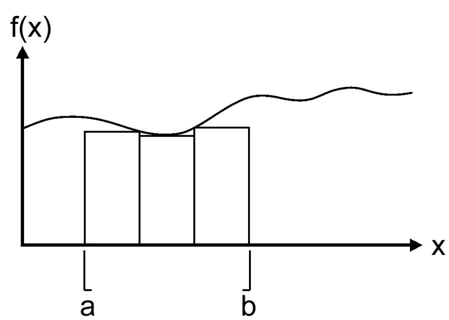
1 Introduction

Measure theory serves as the mathematical foundation for probability, real analysis, and integration. At its core lies the concept of a σ-algebra, a structured collection of subsets that ensures measurability and enables the rigorous definition of measures.

A σ-algebra provides the necessary framework for defining measurable spaces, allowing for the assignment of measures in a consistent manner. This structure is fundamental in handling infinite processes, such as countable unions and intersections, which play a crucial role in probability theory and real analysis.

The development of Lebesgue integration builds upon the concept of σ-algebras, addressing limitations in the Riemann integral by allowing the integration of a broader class of functions. These ideas are essential in understanding measurable functions, probability spaces, and advanced mathematical tools used in modern applications, including artificial intelligence and statistical modeling.

2 Riemann Integral

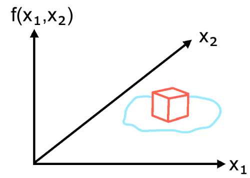


The **Riemann integral** sums function values over subintervals of a domain to approximate the area under the curve. It partitions the domain [a,b] into subintervals and sums up function values multiplied by interval widths. The Riemann integral is defined as:

$$\int_a^b f\,dt \approx \sum_k \mathrm{vol}(I_k) \cdot f(m_k), \quad \text{where } \mathrm{vol}(I_k) = x_{k+1} - x_k. \tag{1}$$

2.1 Problems with Riemann Integral

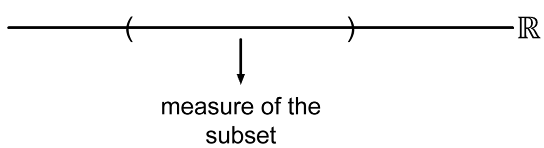
- **Difficult to extend to higher dimensions:** Partitioning becomes more complicated in higher dimensions since integration involves summing over volumes.



- **Dependence on continuity:** If a function has too many discontinuities, the function values fluctuate wildly, making it impossible to approximate the area under the curve with sums.
- **Limit processes issue:** The interchange of limit and integration is not always valid in Riemann integration. If the sequence of functions $f_n(x)$ does not converge uniformly, the two sides of the equation may yield different results.

$$\lim_{n\to\infty} \int_a^b f_n(x)dx \stackrel{?}{=} \int_a^b \lim_{n\to\infty} f_n(x)dx$$

3 Lebesgue Integrals



The **Lebesgue integral** generalizes the Riemann integral by measuring the contribution of function values based on the **measure of the subset** where they occur, rather than summing over fixed partitions of the domain.

Instead of dividing the x-axis into intervals (as in Riemann integration), Lebesgue integration groups together points where the function takes similar values and integrates over those sets.

For example, if f is a constant function over a measurable subset A, its integral is computed as:

$$\int_A f\,d\mu = c \cdot \mu(A), \quad \text{if } f(x) = c \text{ for all } x \in A.$$

This perspective allows us to integrate **more general functions**, including those with infinitely many discontinuities (e.g., the indicator function of rationals in $[0, 1]$). However, for a function f to be Lebesgue integrable, it must be **measurable** and satisfy:

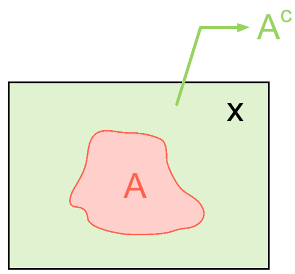
$$\int |f|\,d\mu < \infty. \tag{2}$$

The Lebesgue integral is particularly useful in probability theory, functional analysis, and modern applications of machine learning, where handling measure-theoretic probability spaces is essential.

4 Definition of σ-Algebra

Definition 1 *Let X be a set, and let $\mathcal{P}(X)$ be its power set. A collection of subsets $\mathcal{A} \subseteq \mathcal{P}(X)$ is called a* ***σ-algebra*** *if:*

- $\emptyset, X \in \mathcal{A}$.
- *If $A \in \mathcal{A}$, then $A^c := X \setminus A \in \mathcal{A}$.*
- *If $A_i \in \mathcal{A}$ for all $i \in \mathbb{N}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.*



4.1 Role in Probability Theory

A σ-algebra is a collection of subsets closed under countable set operations, providing the structure needed to define a probability measure. This allows for the formal and consistent assignment of probabilities to events, which is essential for solving numerous real-world problems.

4.2 Example: Probability Spaces

In a probability space, the σ-algebra represents all possible events for which we can assign probabilities. For instance, consider a fair six-sided die roll, where the sample space is:

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

The power set $\mathcal{P}(\Omega)$ is the largest possible σ-algebra, containing all subsets of Ω, but a more practical σ-algebra might include only meaningful events such as:

$$\mathcal{A} = \{\emptyset, \Omega, \{1, 3, 5\}, \{2, 4, 6\}\}.$$

This σ-algebra allows us to assign probabilities, such as:

$$P(\{1, 3, 5\}) = \frac{3}{6}, \quad P(\{2, 4, 6\}) = \frac{3}{6}.$$

Understanding the structure of σ-algebras ensures that probability assignments remain consistent.

4.3 Applications in Machine Learning

σ-algebras play a critical role in defining probability spaces, which are essential for modeling uncertainty and randomness in machine learning. In probabilistic graphical models, σ-algebras formalize dependencies between random variables, enabling the computation of joint and conditional probabilities. This structure is crucial for algorithms that perform inference and learning in complex models.

5 Measurable Spaces

Definition 2 *A **measurable space** is a set X with a σ-algebra $\mathcal{A}$ over X. It is denoted as $(X, \mathcal{A})$, and the sets in $\mathcal{A}$ are called **$\mathcal{A}$-measurable sets**.*

5.1 Examples

- **Smallest σ-algebra:** $\{\emptyset, X\}$.
- **Largest σ-algebra:** Power set $\mathcal{P}(X)$.

6 Intersection of σ-Algebras

Definition 3 *Let $\mathcal{A}_i$ be σ-algebras on X, indexed by $i \in I$. Then their intersection*

$$\bigcap_{i \in I} \mathcal{A}_i \tag{3}$$

is also a σ-algebra on X.

7 Generated σ-Algebra

Definition 4 *For $\mathcal{M} \subseteq \mathcal{P}(X)$, there is a smallest σ-algebra that contains $\mathcal{M}$. This is called the σ-algebra generated by $\mathcal{M}$:*

$$\sigma(\mathcal{M}) := \bigcap \{\mathcal{A} \mid \mathcal{M} \subseteq \mathcal{A}, \mathcal{A} \text{ is a } \sigma\text{-algebra}\}. \tag{4}$$

7.1 Example

Let $X = \{a, b, c, d\}$ and $\mathcal{M} = \{\{a\}, \{b\}\}$. Then the generated σ-algebra is:

$$\sigma(\mathcal{M}) = \{\emptyset, X, \{a\}, \{b\}, \{a, b\}, \{b, c, d\}, \{a, c, d\}, \{c, d\}\}.$$

8 Borel σ-Algebra

Definition 5 *Let (X, τ) be a **topological space**. This means:*

- *X can be a **metric space**, meaning it has a notion of distance.*
- *X can be a subset of $\mathbb{R}^n$.*
- *The collection of **open sets** in X defines the topology τ.*

*The **Borel σ-algebra** on X, denoted as $B(X)$, is the σ-algebra generated by the open sets of X:*

$$B(X) := \sigma(\tau). \tag{5}$$

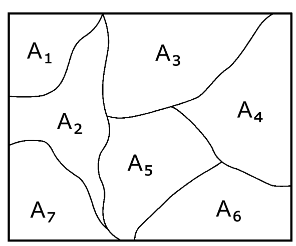
9 Measures

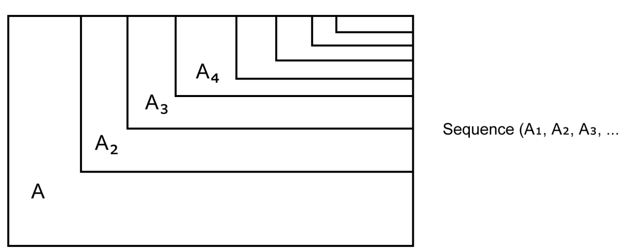
Definition 6 *Let $(X, \mathcal{A})$ be a measurable space. A function $\mu : \mathcal{A} \to [0, \infty]$ is called a **measure** if it satisfies:*

- *(a) $\mu(\emptyset) = 0$.*
- *(b) (Countable Additivity) For any sequence of pairwise disjoint sets $A_i \in \mathcal{A}$, we have:*

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i), \quad \text{where } A_i \cap A_j = \emptyset \text{ for } i \neq j. \tag{6}$$

This property ensures that measures assign values consistently over countable unions.





Definition 7 *A **measure space** is a measurable space $(X, \mathcal{A})$ equipped with a measure μ. It is denoted as $(X, \mathcal{A}, \mu)$.*

9.1 Examples

- (a) **Counting measure:** Assigns a value equal to the number of elements in A, or infinity if A is infinite. It is useful in discrete probability spaces, such as analyzing datasets where each element is counted.

  If $X, \mathcal{A} = \mathcal{P}(X)$, define $\mu(A)$ as:

$$\mu(A) = \begin{cases} \#A, & \text{if } A \text{ has finitely many elements} \\ \infty, & \text{else} \end{cases} \tag{7}$$

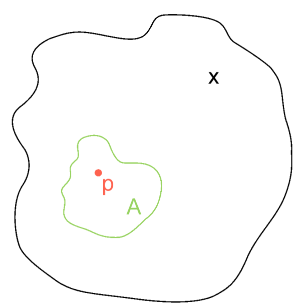
  Calculation rules in $[0, \infty]$:

  - $x + \infty = \infty$ for all $x \in [0, \infty]$.
  - $x \cdot \infty = \infty$ for all $x \in [0, \infty]$.
  - $0 \cdot \infty = 0$ (in most cases in measure theory).

- (b) **Dirac measure:** Assigns measure 1 to a single point p and 0 elsewhere, behaving like a "spike" of probability at p. In machine learning, Dirac measures appear in reinforcement learning, where deterministic policies can be represented using Dirac delta functions.

  Given $p \in X$, define $\delta_p(A)$ as:

$$\delta_p(A) = \begin{cases} 1, & p \in A \\ 0, & \text{else} \end{cases} \tag{8}$$

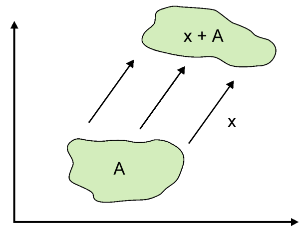


- (c) **Lebesgue measure:** Extends the concept of length, area, and volume in $\mathbb{R}^n$, ensuring translation invariance. Many probability distributions (e.g., Gaussian, Exponential) are defined using Lebesgue measures, allowing integration over continuous spaces.

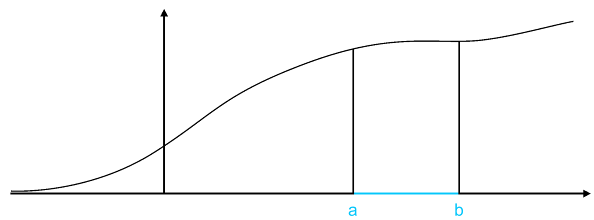
  Define a measure on $X = \mathbb{R}^n$ such that:

$$\mu([0,1]^n) = 1, \quad \mu(x + A) = \mu(A) \quad \forall x \in \mathbb{R}^n. \tag{9}$$

  (σ-algebra $\neq$ the power set)

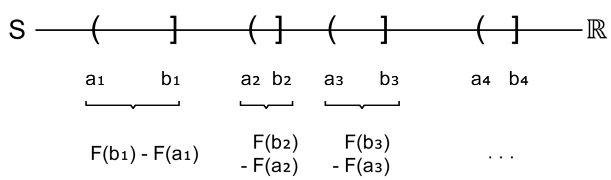


- (d) **A More Useful Class of Measures:** Let $X = \mathbb{R}$, and let $\mathcal{A}$ be the Borel σ-algebra. Let $F : \mathbb{R} \to \mathbb{R}$ is a **monotonically increasing, continuous function**.



Define a measure μ_F on $(\mathbb{R}, \mathcal{A})$ by setting:

$$\mu_F(S) = \inf \left\{ \sum_{j=1}^{\infty} (F(b_j) - F(a_j)) \mid S \subseteq \bigcup_{j=1}^{\infty} (a_j, b_j] \right\}. \tag{10}$$



This measure is defined by covering S using intervals and assigning an "elementary volume" $F(b) - F(a)$ to each interval. The best covering is taken using the infimum.

10 Null Sets and Almost Everywhere

Definition 8 *A subset $N \in \mathcal{A}$ is called a* ***null set*** *if $\mu(N) = 0$. A property holds* ***almost everywhere*** *if it holds for all $x \in X$ except for x in a null set N. In probability theory, this is referred to as* ***almost surely****.*

References

[1] σ-algebra. Available at: https://en.wikipedia.org/wiki/%CE%A3-algebra. Accessed March 2025.

[2] Lebesgue Integral. Available at: https://en.wikipedia.org/wiki/Lebesgue_integral#. Accessed March 2025.

[3] Sigma Algebra. Available at: https://www.geeksforgeeks.org/sigma-algebra/. Accessed March 2025.

[4] Probability Measures. Kyle Siegrist. Available at: https://stats.libretexts.org/Bookshelves/Probability_Theory/Probability_Mathematical_Statistics_and_Stochastic_Processes_(Siegrist)/02%3A_Probability_Spaces/2.03%3A_Probability_Measures. Accessed March 2025.

[5] Measure Theory Basics. Will Fithian. Available at: https://www.stat.berkeley.edu/~wfithian/courses/stat210a/measure-theory-basics.html. Accessed March 2025.

[6] Reinforcement Learning in Continuous Time and Space: A Stochastic Control Approach. Haoran Wang, Thaleia Zariphopoulou, and Xun Yu Zhou. Available at: https://www.columbia.edu/~xz2574/download/rl.pdf. Accessed March 2025.