Markov chains play an essential role in all five stages of data interpretation, namely

i. Generation,
ii. Compression,
iii. Analysis,
iv. Inference, and
v. Prediction.

Applications include

i. Queuing theory,
ii. Population Biology,
iii. generating test data,
iv. Optimization,
v. counting or more generally approximating volume and integration,
vi. analyzing typical behavior for random objects, and more.

We will review background materials on set theory, measure theory, and linear algebra before delving into the main topic.

1.1 Sequences, Sets, and Sequences of Sets

Given a set $S \subseteq \mathbb{R}$ we define

$$
\sup S := \inf \{ u : \text{for all } x \in S, x \leq u \}
\quad \text{and} \quad
\inf S := \sup \{ \ell : \text{for all } x \in S, x \geq \ell \}.
$$

The well-ordering principle says that infimum and supremum always exist in $[-\infty, \infty]$. For example,

$$
\inf [a, b] = \inf \{a, b\} = \inf \{a, b\} = a, \quad \sup (a, \infty) = \infty, \quad \inf \emptyset = \infty, \quad \sup \emptyset = -\infty, \quad \inf \emptyset = \infty.
$$

Given a sequence of real numbers $a_n, n \geq 1$, we say that $a_n \to L$ or $\lim_{n \to \infty} a_n = L$ for some $L \in \mathbb{R}$ iff

for every $\varepsilon > 0$ there exists $N = N(\varepsilon)$ such that for all $n \geq N$, $|a_n - L| < \varepsilon$.

We say that $a_n \to \infty$ iff for every $M > 0$ there exists $N = N(M)$ such that for all $n > N$, $a_n > M$.

Note that $a_n$ can diverge without going to $\pm \infty$, e.g., by oscillating.
Given a sequence of real numbers \( a_n, n \geq 1 \), we define

\[
\limsup_{n \to \infty} a_n := \lim_{m \to \infty} \sup_{n \geq m} a_n = \inf_{m \geq 1} \sup_{n \geq m} a_n
\]

and

\[
\liminf_{n \to \infty} a_n := \lim_{m \to \infty} \inf_{n \geq m} a_n = \sup_{m \geq 1} \inf_{n \geq m} a_n.
\]

Note that, \( \lim a_n \) exists iff \( \limsup a_n = \liminf a_n \). We also have the following result.

**Theorem 1.1** (Monotone Convergence Theorem – for real numbers). A bounded monotone sequence of real numbers has a limit. If the sequence is increasing, the limit is the supremum, and if the sequence is decreasing, the limit is the infimum of the sequence.

We can extend the above definition to a sequence of sets. Given a sequence of sets \( A_n \subseteq \mathbb{R}, n \geq 1 \) we define

\[
\bigcup_{n \geq 1} A_n := \{ x : \exists \ n \text{ s.t. } x \in A_n \} = A_1 \cup A_2 \cup \cdots \cup A_n \cup \cdots
\]

\[
\bigcap_{n \geq 1} A_n := \{ x : \forall \ n, x \in A_n \} = A_1 \cap A_2 \cap \cdots \cap A_n \cap \cdots
\]

DeMorgan’s Laws say that:

\[
\left( \bigcup_{n \geq 1} A_n \right)^c = \bigcap_{n \geq 1} A_n^c
\]

\[
\left( \bigcap_{n \geq 1} A_n \right)^c = \bigcup_{n \geq 1} A_n^c.
\]

Given a sequence of sets \( A_n, n \geq 1 \), we now define

\[
\limsup_{n \to \infty} A_n := \bigcap_{m \geq 1} \bigcup_{n \geq m} A_n
\]

\[
\liminf_{n \to \infty} A_n := \bigcup_{m \geq 1} \bigcap_{n \geq m} A_n.
\]

Given a set \( E \), we define the power set of \( E \) as

\[
2^E := \{ A \mid A \subseteq E \}
\]

as the collection of all subsets of \( E \). When \( |E| < \infty \), there is a one-to-one correspondence between subsets of \( E \) and binary strings of length \( |E| \), i.e., \( \{0, 1\}^E \).

Note that, \( E \) is finite iff there exists a bijection \( f : E \to \{1, 2, \ldots, n\} \) for some non-negative integer \( n \). Then we say that \( |E| = n \) or \( E \) has \( n \) elements. Similarly, \( E \) is countably infinite iff there exists a bijection \( f : E \to \mathbb{N} := \{1, 2, \ldots\} \). We say a set \( E \) is countable if it is finite or countably infinite. For example, \( \mathbb{R} \) is not countable but \( \mathbb{Q} \) is countable. \( 2^\mathbb{N} \) is also not countable.

### 1.2 σ-algebras and partitions
Definition 1.2. Given a set $E$, a $\sigma$-algebra $\mathcal{E}$ on $E$ is a collection of subsets of $E$ such that

(a) $\emptyset \in \mathcal{E}$
(b) $A \in \mathcal{E} \implies A^c \in \mathcal{E}$
(c) $A_n \in \mathcal{E}$ for all $n \in \mathbb{N} \implies \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{E}$.

The tuple $(E, \mathcal{E})$ is called a measurable space.

Definition 1.3. Given two $\sigma$-algebras $\mathcal{E}_1 \subseteq \mathcal{E}_2$ on $E$, we say that $\mathcal{E}_1$ is coarser than $\mathcal{E}_2$ and $\mathcal{E}_2$ is finer than $\mathcal{E}_1$.

Note that, $2^E$ is the biggest or finest $\sigma$-algebra on $E$ and $\{\emptyset, E\}$ is the smallest or coarsest $\sigma$-algebra on $E$, called the trivial $\sigma$-algebra.

Definition 1.4. If $A$ is a collection of subsets of $E$, i.e., $A \subseteq 2^E$, the $\sigma$-algebra generated by $A$, written as $\sigma(A)$, is the smallest or coarsest $\sigma$-algebra containing $A$.

Definition 1.5. A partition $\Pi$ of $E$ is a collection of subsets of $E$, namely $\Pi = \{E_i | i \in I\}$ that is

(a) Exhaustive: $\bigcup_{i \in I} E_i = E$
(b) Exclusive: $i \neq j \implies E_i \cap E_j = \emptyset$.

Example 1.6. For $E = \{1, 2, 3\}$, $\Pi = \{\{1\}, \{2, 3\}\}$ is a partition of $E$.

Given a partition $\Pi = \{E_i\}_{i \in I}$ of a countable set $E$, one can write down $\sigma(\Pi)$ explicitly as follows:

$$\sigma(\Pi) = \{\bigcup_{A \subseteq I} E_i | A \subseteq I\}.$$

We will leave the proof as an exercise. Partitions can be formed as equivalence classes based on equivalence relations and vice versa. Recall that,

Definition 1.7. An equivalence relation $\sim$ on $E$ is

(a) reflexive: $x \sim x \forall x \in E$,
(b) symmetric: $x \sim y \implies y \sim x$ and
(c) transitive: $x \sim y, y \sim z \implies x \sim z$.

An equivalence class containing $x$ is $E_x := \{y \in E | x \sim y\}$. Note that, $\{E_x | x \in E\}$ gives a partition of $E$.

Conversely, given a countable set $E$, any $\sigma$-algebra $\mathcal{F}$ on $E$ arises in the above way, i.e., there exists a partition $\Pi$ on $E$ such that $\mathcal{F} = \sigma(\Pi)$. Define the relation $\sim$ on $E$ as follows: $x \sim y$ if $(x \in A$ iff $y \in A$ for all $A \in \mathcal{F}$). $\sim$ is an equivalence relation. Let $\Pi := \{E_i\}_{i \in I}$ be the disjoint equivalence classes w.r.t. $\sim$ giving a partition of $E$. One can prove that, $\mathcal{F} = \sigma(\Pi)$. If $E$ is uncountable, then the above statement is not true.
1.3 Fubini’s theorem

Theorem 1.8 (Fubini’s theorem). If $I$ and $J$ are countable sets and $a_{ij} \geq 0$ for all $i \in I, j \in J$, then

$$\sum_{i \in I} \sum_{j \in J} a_{ij} = \sum_{j \in J} \sum_{i \in I} a_{ij}.$$ 

That is, we can rearrange the order of summing countable many non-negative terms.

Proof. Enumerate $J = \{j_1, j_2, \ldots \}$. Then

$$\sum_{i \in I} \sum_{j \in J} a_{ij} \geq \sum_{i \in I} \left( \sum_{j=1}^{n} a_{ij} \right)$$

by dropping non-negative terms. Now one sum is finite, so we can exchange order to get

$$\sum_{i \in I} \left( \sum_{j=1}^{n} a_{ij} \right) = \sum_{j=1}^{n} \left( \sum_{i \in I} a_{ij} \right).$$

The last sum forms an increasing sequence in $n$ that converges to $\sum_{j \in J} \sum_{i \in I} a_{ij}$ by monotone convergence theorem. Thus

$$\sum_{i \in I} \sum_{j \in J} a_{ij} \geq \sum_{j \in J} \sum_{i \in I} a_{ij}.$$ 

By symmetry, the other inequality is also true, so equality holds. ■