Markov Chains:
The Chapman-Kolmogorov equation, and stationary distributions
The future of the system is determined entirely by its present, regardless of the past history that led to its present state.
Let us recall the definition of a Markov Chain.

Take a sequence $X_1, X_2, X_3, \ldots$ of random variables. This sequence is said to be a **Markov Chain** if

\[
P(X_{n+1} = j \mid X_n = i, X_{k_1} = \alpha_1, \ldots, X_{k_m} = \alpha_m)
\]

(where the indices $k_1, \ldots, k_m$ are all strictly less than $n$)

is a number entirely determined by $n$, $i$, and $j$. This number is called **the transition probability** from $i$ to $j$ at time $n$. 
The Markov Property

Transition probability matrix

If the number is also independent of $n$, then we say the Markov chain is **homogeneous** and the transition probability is denoted simply $p(i, j)$.

*The $i, j$ and the $\alpha_k$’s here denote possible states of the system, they could be numbers, or any other kind of label for the states.*

For convenience, let us think of the states as being labeled from 1 through $N$, so that $i, j = 1, \ldots, N$.

We may think of the transition probabilities $p(i, j)$ as arranged along rows and columns corresponding to $i$ and $j$. 
One more thing: given any $i$, we always have

$$\sum_j p(i, j) = 1$$

**Exercise:** Think about why this must be so!
Examples of Markov Chains

Example 1: (Gambler’s Ruin with $M = 5$)

This is a system with 6 possible states, labeled 0 through 5. If the coin is fair, the probabilities are

\[
\begin{array}{cccccc}
0 & 1 & 2 & 3 & 4 & 5 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0.5 & 0 & 0.5 & 0 & 0 & 0 \\
2 & 0 & 0.5 & 0 & 0.5 & 0 & 0 \\
3 & 0 & 0 & 0.5 & 0 & 0.5 & 0 \\
4 & 0 & 0 & 0 & 0.5 & 0 & 0.5 \\
5 & 0 & 0 & 0 & 0 & 0 & 1 \\
\end{array}
\]
Examples of Markov Chains

Example 1: (Gambler’s Ruin with $M = 5$)

This arrangement is no accident.
It’s natural—and extremely useful—to think of the transition probabilities as forming a $N \times N$ matrix

$$
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.5 & 0 & 0.5 & 0 & 0 & 0 & 0 \\
0 & 0.5 & 0 & 0.5 & 0 & 0 & 0 \\
0 & 0 & 0.5 & 0 & 0.5 & 0 & 0 \\
0 & 0 & 0 & 0.5 & 0 & 0.5 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 \\
\end{pmatrix}
$$

This is called the transition matrix for the chain.
Examples of Markov Chains

**Example 1:** (Gambler’s Ruin with $M = 5$)

All the **essential information about the Markov chain** is contained in the transition matrix.

\[
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.5 & 0 & 0.5 & 0 & 0 & 0 & 0 \\
0 & 0.5 & 0 & 0.5 & 0 & 0 & 0 \\
0 & 0 & 0.5 & 0 & 0.5 & 0 & 0 \\
0 & 0 & 0 & 0.5 & 0 & 0.5 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\]
Examples of Markov Chains

Example 1: (Gambler’s Ruin with $M = 5$)

If the coin were biased, we would have two numbers $p, q \in [0, 1]$ with $p + q = 1$ $p \neq q$, and the transition matrix is

\[
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
p & 0 & q & 0 & 0 & 0 & 0 \\
0 & p & 0 & q & 0 & 0 & 0 \\
0 & 0 & p & 0 & q & 0 & 0 \\
0 & 0 & 0 & p & 0 & q & 0 \\
0 & 0 & 0 & 0 & p & 0 & q \\
0 & 0 & 0 & 0 & 0 & p & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\]
Examples of Markov Chains

Example 2: (a two-state model for weather)

Denote 1 = Rainy Day, 2 = Sunny Day.

\[
\begin{array}{cc}
1 & 2 \\
1 & 0.6 & 0.4 \\
2 & 0.2 & 0.8 \\
\end{array}
\]

For Rainy days:
Next day will be Rainy or Sunny with probabilities 0.6 and 0.4.

For Sunny days:
Next day will be Rainy or Sunny with probabilities 0.2 and 0.8.
Example 2: (a two-state model for weather)

Again, we arrange these probabilities in a matrix,

\[
\begin{pmatrix}
0.6 & 0.4 \\
0.2 & 0.8
\end{pmatrix}
\]

Since the system only has two states, the matrix is $2 \times 2$.

Question: Out of $N$ days, what proportion of them are rainy?
Examples of Markov Chains

Example 2: (a two-state model for weather)

\[
\begin{pmatrix}
0.6 & 0.4 \\
0.2 & 0.8
\end{pmatrix}
\]

Fix \( i \) and \( j \), what is the probability of going from \( i \) to \( j \) in two steps? i.e. what is \( \Pr(X_2 = j \mid X_0 = i) \)?
Examples of Markov Chains

Example 2: (a two-state model for weather)

\[
\begin{pmatrix}
0.6 & 0.4 \\
0.2 & 0.8
\end{pmatrix}
\]

Well, there are two ways to go from $i$ to $j$ in exactly two steps,

\[
i \mapsto 1 \mapsto j
\]
\[
i \mapsto 2 \mapsto j
\]

Then, $P(X_2 = j \mid X_0 = i)$ is equal to

\[
P(X_2 = j, X_1 = 1 \mid X_0 = i) + P(X_2 = j, X_1 = 2 \mid X_0 = i)
\]
Examples of Markov Chains

Example 2: (a two-state model for weather)

\[
\begin{pmatrix}
0.6 & 0.4 \\
0.2 & 0.8
\end{pmatrix}
\]

In terms of conditional probabilities, this sum is equal to

\[
P(X_2 = j \mid X_0 = i, X_1 = 1)P(X_1 = 1 \mid X_0 = i) + P(X_2 = j \mid X_0 = i, X_1 = 2)P(X_1 = 2 \mid X_0 = i)
\]

Then, the Markov property says this is equal to

\[
P(X_2 = j \mid X_1 = 1)P(X_1 = 1 \mid X_0 = i) + P(X_2 = j \mid X_1 = 2)P(X_1 = 2 \mid X_0 = i)
\]
Example 2: (a two-state model for weather)

\[
\begin{pmatrix}
0.6 & 0.4 \\
0.2 & 0.8
\end{pmatrix}
\]

In other words, \( P(X_2 = j \mid X_0 = i) \) is given by

\[
p(1, j)p(i, i) + p(2, j)p(i, 2)
\]
Examples of Markov Chains

Example 2: (a two-state model for weather)

\[
\begin{pmatrix}
0.6 & 0.4 \\
0.2 & 0.8
\end{pmatrix}
\]

Let us write \( p^2(i, j) = P(X_{n+2} = j \mid X_2 = i) \), then

\[
p^2(i, j) = p(1, j)p(i, i) + p(2, j)p(i, 2)
\]

These are called the 2-step probabilities. The above formula is exactly the formula for matrix multiplication!
Examples of Markov Chains

**Example 2:** (a two-state model for weather)

\[
\begin{pmatrix}
0.6 & 0.4 \\
0.2 & 0.8 \\
\end{pmatrix}
\]

THEN: 2-step probabilities \( p^2(i, j) \) yield a new \( 2 \times 2 \) matrix, and

\[
\begin{pmatrix}
0.6 & 0.4 \\
0.2 & 0.8 \\
\end{pmatrix}^2 = \begin{pmatrix}
0.44 & 0.56 \\
0.28 & 0.72 \\
\end{pmatrix}
\]

i.e. the matrix of 2-step probabilities \( p^2(i, j) \) is simply the **square** of the matrix of transition probabilities \( p(i, j) \).
Examples of Markov Chains

Example 2: (a two-state model for weather)

Question: Out of \( N \) days, what proportion of them are rainy?

Observe, the adjoint matrix has an interesting eigenvector

\[
\begin{pmatrix}
0.6 & 0.4 \\
0.2 & 0.8
\end{pmatrix}^t \begin{pmatrix}
1/3 \\
2/3
\end{pmatrix} = \begin{pmatrix}
1/3 \\
2/3
\end{pmatrix}
\]

The eigenvector corresponds to a probability distribution!.

Later, one of our theorems will guarantee that (with high prob.)

\( 1/3 \) of the \( N \) days are rainy

\ldots at least, when \( N \) is large.
If $X_1, X_2, \ldots$ denotes a homogeneous Markov chain, then the $n$-step transition probabilities are defined by

$$p^n(i, j) = P(X_n = j \mid X_0 = i)$$
n-step transition probabilities

If $X_1, X_2, \ldots$ denotes a homogeneous Markov chain, then the $n$-step transition probabilities are defined by

$$p^n(i, j) = P(X_n = j \mid X_0 = i)$$

Since the chain is homogeneous, for any $k$ we have

$$P(X_{k+n} = j \mid X_k = i) = p^n(i, j)$$

Accordingly, $p^n(i, j)$ defines a $N \times N$ matrix (where $N$ is the number of states in the system), this is called the $n$-step transition matrix, which we may denote simply as $p^n$. 
Theorem (Chapman-Kolmogorov)

Let $n, m \geq 0$, then, in terms of matrix multiplication, we have

$$p^{n+m} = p^n p^m$$

In particular, the matrix $p^n$ is simply the $n$th-power of the transition probability matrix.
The total probability formula says that

\[
P(X_{n+1} = j) = \sum_k P(X_{n+1} = j \mid X_n = k)P(X_n = k)
\]

\[
= \sum_k p(k, j)P(X_n = k)
\]

This has a linear algebra interpretation: if we introduce a transition probability matrix with entries \( p = (p(i, j))_{ij} \), then to go from the distribution of \( X_n \) to the distribution of \( X_{n+1} \), all we need to do is multiply the “vector” of probabilities by the adjoint of the matrix \( p \).
...all we need to do is multiply the “vector” of probabilities by the **adjoint** of the matrix $p$.

This observation leads to what is known as the **Chapman-Kolmogorov** equation, which allows us to compute probabilities of the state several time steps into the future.