4 Markov Chains

4.1 Basic Notions
4.2 Irreducibility
4.3 Transience/Recurrence
4.4 Invariant Measures
4.5 Ergodicity and stationarity
4.6 Limit Theorems
Use of Markov chains

- Many algorithms can be described as Markov chains

Needed properties

- The quantity of interest is what the chain converges to

We need to know

- When will chains converge
- What do they converge to
4.1 Basic notions

A *Markov chain* is a sequence of random variables that can be thought of as evolving over time.

Probability of a transition depends on the particular set that the chain is in

Chain defined through its *transition kernel*

A *transition kernel* is a function $K$ defined on $\mathcal{X} \times \mathcal{B}(\mathcal{X})$ such that

(i). $\forall x \in \mathcal{X}, K(x, \cdot)$ is a probability measure;

(ii). $\forall A \in \mathcal{B}(\mathcal{X}), K(\cdot, A)$ is measurable.
• When $\mathcal{X}$ is *discrete*, the transition kernel simply is a (transition) matrix $K$ with elements

$$P_{xy} = P(X_n = y | X_{n-1} = x), \quad x, y \in \mathcal{X}.$$

• In the continuous case, the *kernel* also denotes the conditional density $K(x, x')$ of the transition $K(x, \cdot)$

$$P(X \in A | x) = \int_A K(x, x') dx'.$$
Given a transition kernel $K$, a sequence $X_0, X_1, \ldots, X_n, \ldots$ of random variables is a Markov chain denoted by $(X_n)$, if, for any $t$, the conditional distribution of $X_t$ given $x_{t-1}, x_{t-2}, \ldots, x_0$ is the same as the distribution of $X_t$ given $x_{t-1}$. That is,

$$P(X_{k+1} \in A| x_0, x_1, x_2, \ldots, x_k) = P(X_{k+1} \in A| x_k)$$

$$= \int_A K(x_k, dx)$$
Example 22 – AR(1) Models –

Simple illustration of Markov chains on continuous state space

\[ X_n = \theta X_{n-1} + \varepsilon_n , \quad \theta \in \mathbb{R}, \]

with \( \varepsilon_n \sim N(0, \sigma^2) \)

If the \( \varepsilon_n \)'s are independent, \( X_n \) independent from \( X_{n-2}, X_{n-3}, \ldots \) conditionally on \( X_{n-1} \).
Chain dependence

Note that the entire structure of the chain only depends on

- The transition function $K$
- The initial state $x_0$ or initial distribution $X_0 \sim \mu$
4.2 Irreducibility

Irreducibility is one measure of the sensitivity of the Markov chain to initial conditions.

It leads to a guarantee of convergence.

In the discrete case, the chain is irreducible if all states communicate, namely if

\[ P_x(\tau_y < \infty) > 0, \quad \forall x, y \in X, \]

\( \tau_y \) being the first time \( y \) is visited.
In the continuous case, the chain is \( \varphi \)-irreducible for some measure \( \varphi \) if for some \( n \),

\[
K^n(x, A) > 0
\]

- for all \( x \in \mathcal{X} \)
- for every \( A \in \mathcal{B}(\mathcal{X}) \) with \( \varphi(A) > 0 \)
Example 23 – AR(1) again –

\[ X_{n+1} = \theta X_n + \varepsilon_{n+1} \]

with \( \varepsilon_n \) iid normal variables,

- The chain is irreducible
- The reference measure \( \varphi \) is Lebesgue measure
- In fact, \( K(x, A) > 0 \) for every \( x \in \mathbb{R} \) and every \( A \) such that \( \lambda(A) > 0 \).
If \( \varepsilon_n \) is uniform on \([-1, 1]\) and \(|\theta| > 1\),

\[
X_{n+1} - X_n \geq (\theta - 1)X_n - 1 \geq 0
\]

for \( X_n \geq 1/(\theta - 1) \), the chain is increasing and cannot visit previous values.
4.2.1 Cycles and Aperiodicity

Sometimes deterministic constraints on the moves from $X_n$ to $X_{n+1}$.

In the discrete case, the *period* of a state $\omega \in \mathcal{X}$ is

$$d(\omega) = \text{g.c.d.} \left\{ m \geq 1; K^m(\omega, \omega) > 0 \right\},$$

where *g.c.d.* is the greatest common denominator.
For an irreducible chain on a finite space $\mathcal{X}$, the transition matrix is a block matrix

$$P = \begin{pmatrix}
0 & D_1 & 0 & \cdots & 0 \\
0 & 0 & D_2 & 0 & \\
& & \ddots & \ddots & \\
D_d & 0 & 0 & 0 & 0
\end{pmatrix},$$

where the blocks $D_i$ are stochastic matrices.

- From block 1 you must go to block 2, from 2 to 3, etc.
- You return to the initial group every $d$-th step
If the chain is irreducible (so all states communicate) only one value for the period. An irreducible chain is *aperiodic* if it has period 1. If one state $x \in \mathcal{X}$ satisfies $P_{xx} > 0$, the chain $(X_n)$ is aperiodic, although this is not a necessary condition.
For continuous chains, similar definition:

- If the transition kernel has density \( f(\cdot|x_n) \), sufficient condition for aperiodicity is that \( f(\cdot|x_n) \) is positive in a neighborhood of \( x_n \) (since the chain can then remain in this neighborhood for an arbitrary number of instants before visiting any set \( A \)).

- For instance, in the AR(1)Example, \( (X_n) \) is aperiodic when \( \varepsilon_n \) is distributed according to \( \mathcal{U}_{[-1,1]} \) and \( |\theta| < 1 \)
4.3 Transience and Recurrence

- Irreducibility ensures that every set $A$ will be visited by the Markov chain $(X_n)$

- This property is too weak to ensure that the trajectory of $(X_n)$ will enter $A$ often enough.

- A Markov chain must enjoy good stability properties to guarantee an acceptable approximation of the simulated model.
  - Formalizing this stability leads to different notions of recurrence
  - For discrete chains, the recurrence of a state equivalent to probability one of sure return.
  - Always satisfied for irreducible chains on finite spaces
In a finite state space $\mathcal{X}$, denote the average number of visits to a state $\omega$ by
\[ \eta_\omega = \sum_{i=1}^{\infty} I_\omega(X_i) \]

If $\mathbb{E}_\omega[\eta_\omega] = \infty$ the state is *recurrent*

If $\mathbb{E}_\omega[\eta_\omega] < \infty$ the state is *transient*

For irreducible chains, recurrence/transience property of the chain, not of a particular state

Similar definitions for the continuous case.
Stronger form of recurrence: **Harris recurrence**

A set $A$ is **Harris recurrent** if

$$P_x(\eta_A = \infty) = 1 \text{ for all } x \in A.$$  

The chain $(X_n)$ is **Harris recurrent** if it is

- $\psi$-irreducible
- for every set $A$ with $\psi(A) > 0$, $A$ is Harris recurrent.

Note that

$$P_x(\eta_A = \infty) = 1 \text{ implies } \mathbb{E}_x[\eta_A] = \infty$$
4.4 Invariant Measures

Stability increases for the chain \( (X_n) \) if marginal distribution of \( X_n \) independent of \( n \)

Requires the existence of a probability distribution \( \pi \) such that

\[
X_{n+1} \sim \pi \quad \text{if} \quad X_n \sim \pi
\]

A measure \( \pi \) is \textbf{invariant} for the transition kernel \( K(\cdot, \cdot) \) if

\[
\pi(B) = \int_X K(x, B) \pi(dx) , \quad \forall B \in \mathcal{B}(X).
\]
Invariant measures

- The chain is **positive recurrent** if \( \pi \) is a probability measure.
- Otherwise it is **null recurrent**
  - If \( \pi \) probability measure also called *stationary distribution* since
    \[
    X_0 \sim \pi \text{ implies that } X_n \sim \pi \text{ for every } n
    \]
  - The stationary distribution is unique
Example 24 – Back to AR(1) –

For the AR(1) model

\[ X_n = \theta X_{n-1} + \varepsilon_n, \quad \theta \in \mathbb{R}, \]

with \( \varepsilon_n \sim N(0, \sigma^2) \), the transition kernel is

\[ \mathcal{N}(\theta x_{n-1}, \sigma^2) \]

and \( \mathcal{N}(\mu, \tau^2) \) is stationary only if

\[ \mu - \theta \mu \quad \text{and} \quad \tau^2 - \tau^2 \theta^2 + \sigma^2. \]
These conditions imply that

\[ \mu = 0, \quad \tau^2 = \sigma^2/(1 - \theta^2), \]

and hence \(|\theta| < 1\).

\(N(0, \sigma^2/(1 - \theta^2))\) is the unique stationary distribution
Paths of a bivariate AR \(-1\) process with \(b = .3\) (recurrent) for 10, 25, 100, and 500 steps. Scale is \(-3\) to 3.

\[\sigma = 1\]
Paths of a bivariate AR-1 process with $b=1.05$ (transient) for 10, 25, 100, and 500 steps. Scale is -20 to 20.
4.5 Ergodicity and convergence

We finally consider: *to what is the chain converging?*

The invariant distribution \( \pi \) natural candidate for the *limiting distribution*

A fundamental property is *ergodicity*, or independence of initial conditions.

In the discrete case, a state \( \omega \) is *ergodic* if

\[
\lim_{n \to \infty} |K^n(\omega, \omega) - \pi(\omega)| = 0.
\]
In general, we establish convergence using the total variation norm

$$\|\mu_1 - \mu_2\|_{TV} = \sup_A |\mu_1(A) - \mu_2(A)|.$$ 

and we want

$$\left\| \int K^n(x, \cdot) \mu(dx) - \pi \right\|_{TV}$$

$$= \sup_A \left| \int K^n(x, A) \mu(dx) - \pi(A) \right|$$

to be small.
If \( (X_n) \) Harris positive recurrent and aperiodic, then

\[
\lim_{n \to \infty} \left\| \int K^n(x, \cdot) \mu(dx) - \pi \right\|_{TV} = 0
\]

for every initial distribution \( \mu \).

We thus take “Harris positive recurrent and aperiodic” as equivalent to “ergodic”

Convergence in total variation implies

\[
\lim_{n \to \infty} \left| \mathbb{E}_\mu[h(X_n)] - \mathbb{E}^\pi[h(X)] \right| = 0
\]

for every bounded function \( h \).
There are different speeds of convergence

- ergodic (fast)
- geometrically ergodic (faster)
- uniformly ergodic (fastest)
4.6 Limit theorems

Ergodicity determines the probabilistic properties of average behavior of the chain.

But also need of statistical inference, made by induction from the observed sample.

If $||P^n_x - \pi||$ close to 0, no direct information about

$$X_n \sim P^n_x$$

We need LLN’s and CLT’s!!!
Classical LLN’s and CLT’s not directly applicable due to:

○ Markovian dependence structure between the observations $X_i$

○ Non-stationarity of the sequence
Ergodic Theorem - LLN

If the Markov chain $(X_n)$ is Harris recurrent, then for any function $h$ with $\mathbb{E}|h| < \infty$, 

$$\lim_{n \to \infty} \frac{1}{n} h(X_i) = \int h(x) d\pi(x),$$
• Recall the stationary distribution of the AR 1 is $\mathcal{N}(0, \sigma^2/(1-\theta^2))$
• For $\sigma = 1$ and $\theta = .5$, $\mathcal{N}(0, (1.15)^2)$
To get a CLT, we need more assumptions. For MCMC, the easiest is reversibility:

A Markov chain \((X_n)\) is reversible if for all \(n\)

\[ X_{n+1}|X_{n+2} \sim X_{n+1}|X_n. \]

The direction of time does not matter
If the Markov chain \((X_n)\) is Harris recurrent and reversible,

\[
\frac{1}{\sqrt{N}} \left( \sum_{n=1}^{N} (h(X_n) - \mathbb{E}^{\pi}[h]) \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \gamma^2_h).
\]

where

\[
0 < \gamma^2_h = \mathbb{E}_\pi[h^2(X_0)] \\
+ 2 \sum_{k=1}^{\infty} \mathbb{E}_\pi[h(X_0)h(X_k)] < +\infty.
\]