Lecture 11: Markov Chains

Deniz Akyildiz

MATH60047/70047 – Stochastic Simulation

November 14, 2022

Imperial College London

Announcements

Announcements

- You can now submit your assignments until 16th of November 2022 through the UG office.
- Solutions will be posted after that date.

A brief recap of what happened:

A brief recap of what happened:

- Sampling methods: $X_i \sim p_\star$ (from now on we will call it $p_\star$)
  - Direct sampling methods (Inversion, transformation)
  - Rejection sampling

A brief recap of what happened:

- Sampling methods: $X_i \sim p_\star$ (from now on we will call it $p_\star$)
  - Direct sampling methods (Inversion, transformation)
  - Rejection sampling
- Integration: $\bar{\varphi} = \int \varphi(x) p_\star(x) \mathrm{d}x$:
  - Using i.i.d samples from $p_\star$:

$$\bar{\varphi} \approx \hat{\varphi}^N_{\mathsf{MC}} = \frac{1}{N} \sum_{i=1}^N \varphi(X_i).$$

  - Using samples from a *proposal* q:

$$\bar{\varphi} \approx \hat{\varphi}^N_{\mathsf{IS}} = \frac{1}{N} \sum_{i=1}^N \bar{\mathsf{w}}_i \varphi(X_i).$$

Today, we will talk about Markov chains.

But why?

But why?

- Recall we are interested in sampling from $p_\star$

But why?

- Recall we are interested in sampling from $p_\star$
- Rejection sampling has its problems (we discussed)

But why?

- Recall we are interested in sampling from $p_\star$
- Rejection sampling has its problems (we discussed)
- Importance sampling is primarily an integration technique, not a sampling technique

But why?

- Recall we are interested in sampling from $p_\star$
- Rejection sampling has its problems (we discussed)
- Importance sampling is primarily an integration technique, not a sampling technique

Yes but how can we use Markov chains?

But why?

- Recall we are interested in sampling from $p_\star$
- Rejection sampling has its problems (we discussed)
- Importance sampling is primarily an integration technique, not a sampling technique

Yes but how can we use Markov chains?

- Markov chains have *stationary* distributions

But why?

- Recall we are interested in sampling from $p_\star$
- Rejection sampling has its problems (we discussed)
- Importance sampling is primarily an integration technique, not a sampling technique

Yes but how can we use Markov chains?

- Markov chains have *stationary* distributions
- We design the chain so that the stationary distribution is $p_\star$!

But why?

- Recall we are interested in sampling from $p_\star$
- Rejection sampling has its problems (we discussed)
- Importance sampling is primarily an integration technique, not a sampling technique

Yes but how can we use Markov chains?

- Markov chains have *stationary* distributions
- We design the chain so that the stationary distribution is $p_\star$!

We will first see how to simulate Markov chains.

What is a Markov chain?

A discrete-time Markov chain is a sequence of random variables $X_0, X_1, \ldots$ such that:

What is a Markov chain?

A discrete-time Markov chain is a sequence of random variables $X_0, X_1, \ldots$ such that:

- X_t depends only on X_{t-1}

What is a Markov chain?

A discrete-time Markov chain is a sequence of random variables $X_0, X_1, \ldots$ such that:

- X_t depends only on X_{t-1}
- In other words, X_t is conditionally independent of $X_0, \ldots, X_{t-2}$ given X_{t-1}.

What is a Markov chain?

A discrete-time Markov chain is a sequence of random variables $X_0, X_1, \ldots$ such that:

- X_t depends only on X_{t-1}
- In other words, X_t is conditionally independent of $X_0, \ldots, X_{t-2}$ given X_{t-1}.

The evolution of the chain is governed by:

What is a Markov chain?

A discrete-time Markov chain is a sequence of random variables $X_0, X_1, \dots$ such that:

- X_t depends only on X_{t-1}
- In other words, X_t is conditionally independent of $X_0, \dots, X_{t-2}$ given X_{t-1}.

The evolution of the chain is governed by:

- A transition matrix M (discrete case)

What is a Markov chain?

A discrete-time Markov chain is a sequence of random variables $X_0, X_1, \ldots$ such that:

- X_t depends only on X_{t-1}
- In other words, X_t is conditionally independent of $X_0, \ldots, X_{t-2}$ given X_{t-1}.

The evolution of the chain is governed by:

- A transition matrix M (discrete case)
- A transition kernel K (continuous case)

What is a Markov chain?

A discrete-time Markov chain is a sequence of random variables $X_0, X_1, \ldots$ such that:

- X_t depends only on X_{t-1}
- In other words, X_t is conditionally independent of $X_0, \ldots, X_{t-2}$ given X_{t-1}.

The evolution of the chain is governed by:

- A transition matrix M (discrete case)
- A transition kernel K (continuous case)

Let us denote our state-space with X.

What is a Markov chain?

The discrete case

A discrete-time Markov chain is a process $(X_n)_{n \in \mathbb{N}}$, when X is finite, satisfies:

What is a Markov chain?

The discrete case

A discrete-time Markov chain is a process $(X_n)_{n\in\mathbb{N}}$, when X is finite, satisfies:

$$\mathbb{P}(X_n = j | X_0 = i_0, \ldots, X_{n-1} = i_{n-1}) = \mathbb{P}(X_n = j | X_{n-1} = i_{n-1}).$$

A discrete-time Markov chain is a process $(X_n)_{n\in\mathbb{N}}$, when X is finite, satisfies:

$$\mathbb{P}(X_n = j | X_0 = i_0, \ldots, X_{n-1} = i_{n-1}) = \mathbb{P}(X_n = j | X_{n-1} = i_{n-1}).$$

We note that *transition matrix* is nothing but these transition probabilities:

$$M_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i).$$

What is a Markov chain?

The discrete case

Imperial College London

A discrete-time Markov chain is a process $(X_n)_{n\in\mathbb{N}}$, when X is finite, satisfies:

$$\mathbb{P}(X_n = j | X_0 = i_0, \ldots, X_{n-1} = i_{n-1}) = \mathbb{P}(X_n = j | X_{n-1} = i_{n-1}).$$

We note that *transition matrix* is nothing but these transition probabilities:

$$M_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i).$$

$$X_0 \to X_1 \to X_2 \to \cdots \to X_n$$

We note that we will only consider the time-homogeneous case, i.e. the transition matrix is time-independent:

$$M_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i) = \mathbb{P}(X_{n-1} = j | X_{n-2} = i) = \cdots$$

We note that we will only consider the time-homogeneous case, i.e. the transition matrix is time-independent:

$$M_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i) = \mathbb{P}(X_{n-1} = j | X_{n-2} = i) = \cdots$$

A Markov chain therefore can be defined entirely by its:

- Initial state (or initial distribution)
- Transition matrix

Let us denote the initial distribution as p_0. This means,

$$p_0(i) = \mathbb{P}(X_0 = i).$$

Let us denote the initial distribution as p_0. This means,

$$p_0(i) = \mathbb{P}(X_0 = i).$$

We consider p_0 as a vector (discrete distribution on support X).

Let us denote the initial distribution as p_0. This means,

$$p_0(i) = \mathbb{P}(X_0 = i).$$

We consider p_0 as a vector (discrete distribution on support X).

Note that in the continuous-space case, we will use the same notation, but we will consider p_0 as a density function.

The transition matrix M is a matrix of size $d \times d$ where $d = |\mathsf{X}|$.

What is a Markov chain?

The discrete case

The transition matrix M is a matrix of size $d \times d$ where $d = |\mathsf{X}|$.

$$M = \begin{bmatrix} M_{11} & M_{12} & \cdots & M_{1d} \\ M_{21} & M_{22} & \cdots & M_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ M_{d1} & M_{d2} & \cdots & M_{dd} \end{bmatrix}.$$

The transition matrix M is a matrix of size $d \times d$ where $d = |\mathsf{X}|$.

$$M = \begin{bmatrix} M_{11} & M_{12} & \cdots & M_{1d} \\ M_{21} & M_{22} & \cdots & M_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ M_{d1} & M_{d2} & \cdots & M_{dd} \end{bmatrix}.$$

We note that this matrix is stochastic, i.e. each row sums to 1:

$$\sum_{j=1}^{d} M_{ij} = 1,$$

since $M_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i)$ and

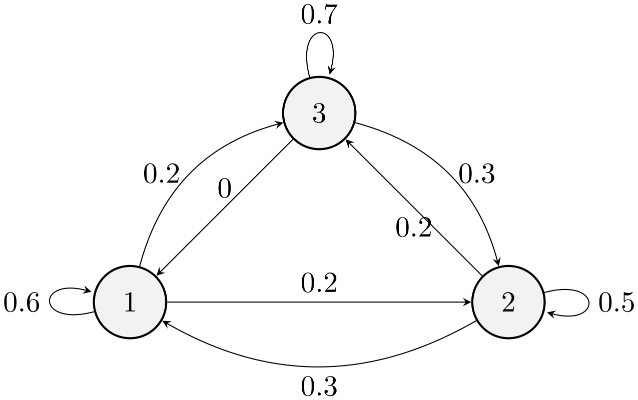
$$\sum_{j=1}^{d} \mathbb{P}(X_{n+1} = j | X_n = i) = 1.$$

What is a Markov chain?

Example 1: Simulate a discrete Markov chain

Consider the transition matrix:

$$M = \begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.3 & 0.5 & 0.2 \\ 0 & 0.3 & 0.7 \end{bmatrix}, \qquad \text{where } \mathsf{X} = \{1, 2, 3\}.$$



M	$X_t = 1$	$X_t = 2$	$X_t = 3$
$X_{t-1} = 1$	0.6	0.2	0.2
$X_{t-1} = 2$	0.3	0.5	0.2
$X_{t-1} = 3$	0	0.3	0.7

Example: Given $X_0 = 1$, how to simulate this chain?

Sample:

$$X_t | X_t = x_{t-1} \sim \text{Discrete}(M_{x_{t-1},\cdot}).$$

What is a Markov chain?

Example 1: Simulate a discrete Markov chain – What does the matrix M mean?

M	$X_t = 1$	$X_t = 2$	$X_t = 3$
$X_{t-1} = 1$	0.6	0.2	0.2
$X_{t-1} = 2$	0.3	0.5	0.2
$X_{t-1} = 3$	0	0.3	0.7

Example: Given $X_0 = 1$, how to simulate this chain?

Sample:

$$X_t | X_t = x_{t-1} \sim \text{Discrete}(M_{x_{t-1},\cdot}).$$

Simulation!

How to compute n-step transition probabilities?

$$M^{(n)} = \mathbb{P}(X_n = j | X_0 = i),$$

where $M^{(n)}$ is a matrix of size $d \times d$.

How to compute n-step transition probabilities?

$$M^{(n)} = \mathbb{P}(X_n = j | X_0 = i),$$

where $M^{(n)}$ is a matrix of size $d \times d$. n-step transition matrix:

$$\begin{aligned}
M_{ij}^{(n)} &= \mathbb{P}(X_n = j | X_0 = i) \\
&= \sum_k \mathbb{P}(X_n = j, X_1 = k | X_0 = i) \\
&= \sum_k \mathbb{P}(X_n = j | X_1 = k, X_0 = i)\mathbb{P}(X_1 = k | X_0 = i) \\
&= \sum_k \mathbb{P}(X_n = j | X_1 = k)\mathbb{P}(X_1 = k | X_0 = i) \\
&= \sum_k M_{ik} M_{kj}^{(n-1)}.
\end{aligned}$$

Therefore, $M^{(n)} = M^n$, nth power.

The Chapman-Kolmogorov equation says that we can obtain

$$\mathbb{P}(X_{n+2} = x_{n+2} | X_n = x_n) =$$
$$\sum_{x_{n+1}} \mathbb{P}(X_{n+2} = x_{n+2} | X_{n+1} = x_{n+1}) \mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n).$$

The Chapman-Kolmogorov equation says that we can obtain

$$\mathbb{P}(X_{n+2} = x_{n+2} | X_n = x_n) = \sum_{x_{n+1}} \mathbb{P}(X_{n+2} = x_{n+2} | X_{n+1} = x_{n+1}) \mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n).$$

This follows from the simple marginalisation rules.

The Chapman-Kolmogorov equation says that we can obtain

$$\mathbb{P}(X_{n+2} = x_{n+2} | X_n = x_n) = \sum_{x_{n+1}} \mathbb{P}(X_{n+2} = x_{n+2} | X_{n+1} = x_{n+1}) \mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n).$$

This follows from the simple marginalisation rules.

However, this gives us a way to relate M^{m+n} to M^m and M^n:

$$\begin{aligned} M^{(m+n)} &= \mathbb{P}(X_{m+n} = j | X_0 = i) \\ &= \sum_k \mathbb{P}(X_{m+n} = j | X_n = k) \mathbb{P}(X_n = k | X_0 = i) \\ &= \sum_k M_{ik}^{(m)} M_{kj}^{(n)}. \end{aligned}$$

Let $p_0(i) = \mathbb{P}(X_0 = i)$ for $i \in \mathsf{X}$.

What is a Markov chain?

The discrete case: The evolution of the density of the chain

Let $p_0(i) = \mathbb{P}(X_0 = i)$ for $i \in \mathsf{X}$. Then, the density of the chain at time n is given by:

$$
\begin{aligned}
p_n(i) &= \mathbb{P}(X_n = i) \\
&= \sum_k \mathbb{P}(X_n = i, X_{n-1} = k) \\
&= \sum_k \mathbb{P}(X_n = i | X_{n-1} = k)\mathbb{P}(X_{n-1} = k) \\
&= \sum_k M_{ki} p_{n-1}(k).
\end{aligned}
$$

What is a Markov chain?

The discrete case: The evolution of the density of the chain

Let $p_0(i) = \mathbb{P}(X_0 = i)$ for $i \in \mathsf{X}$. Then, the density of the chain at time n is given by:

$$\begin{aligned} p_n(i) &= \mathbb{P}(X_n = i) \\ &= \sum_k \mathbb{P}(X_n = i, X_{n-1} = k) \\ &= \sum_k \mathbb{P}(X_n = i | X_{n-1} = k)\mathbb{P}(X_{n-1} = k) \\ &= \sum_k M_{ki} p_{n-1}(k). \end{aligned}$$

This implies that

$$p_n = p_{n-1} M.$$

What is a Markov chain?

The discrete case: The evolution of the density of the chain

Let $p_0(i) = \mathbb{P}(X_0 = i)$ for $i \in \mathsf{X}$. Then, the density of the chain at time n is given by:

$$\begin{aligned} p_n(i) &= \mathbb{P}(X_n = i) \\ &= \sum_k \mathbb{P}(X_n = i, X_{n-1} = k) \\ &= \sum_k \mathbb{P}(X_n = i | X_{n-1} = k)\mathbb{P}(X_{n-1} = k) \\ &= \sum_k M_{ki} p_{n-1}(k). \end{aligned}$$

This implies that

$$p_n = p_{n-1} M.$$

Therefore,

$$p_n = p_0 M^n.$$

Let $p_0(i) = \mathbb{P}(X_0 = i)$ for $i \in \mathsf{X}$. Then, the density of the chain at time n is given by:

$$\begin{aligned} p_n(i) &= \mathbb{P}(X_n = i) \\ &= \sum_k \mathbb{P}(X_n = i, X_{n-1} = k) \\ &= \sum_k \mathbb{P}(X_n = i | X_{n-1} = k)\mathbb{P}(X_{n-1} = k) \\ &= \sum_k M_{ki} p_{n-1}(k). \end{aligned}$$

This implies that

$$p_n = p_{n-1} M.$$

Therefore,

$$p_n = p_0 M^n.$$

Simulation.

Properties of Markov chains

What do we need?

We need Markov chains

Properties of Markov chains

What do we need?

We need Markov chains

- With invariant distributions
- Their convergence is ensured
- Their invariant distribution is unique

We need Markov chains

- With invariant distributions
- Their convergence is ensured
- Their invariant distribution is unique

We will now review the properties which ensure these in discrete space case.

For two states, $x, x' \in \mathsf{X}$, we write $x \rightsquigarrow x'$ if there is a path from x to x':

$$\exists n > 0, \text{ s.t. }, \mathbb{P}(X_n = x' | X_0 = x) > 0.$$

For two states, $x, x' \in \mathsf{X}$, we write $x \rightsquigarrow x'$ if there is a path from x to x':

$$\exists n > 0, \text{ s.t. } , \mathbb{P}(X_n = x' | X_0 = x) > 0.$$

If $x \rightsquigarrow x'$ and $x' \rightsquigarrow x$, then we say that x and x' *communicate*.

For two states, $x, x' \in \mathsf{X}$, we write $x \rightsquigarrow x'$ if there is a path from x to x':

$$\exists n > 0, \text{ s.t. } , \mathbb{P}(X_n = x' | X_0 = x) > 0.$$

If $x \rightsquigarrow x'$ and $x' \rightsquigarrow x$, then we say that x and x' *communicate*.

A communication class $C \subset \mathsf{X}$ is a set of states such that $x \in C$ and $x' \in C$ if and only if $x \rightsquigarrow x'$ and $x' \rightsquigarrow x$.

For two states, $x, x' \in \mathsf{X}$, we write $x \rightsquigarrow x'$ if there is a path from x to x':

$$\exists n > 0, \text{ s.t. } , \mathbb{P}(X_n = x' | X_0 = x) > 0.$$

If $x \rightsquigarrow x'$ and $x' \rightsquigarrow x$, then we say that x and x' *communicate*.

A communication class $C \subset \mathsf{X}$ is a set of states such that $x \in C$ and $x' \in C$ if and only if $x \rightsquigarrow x'$ and $x' \rightsquigarrow x$.

A chain is irreducible if X is a single communication class.

A Markov chain is recurrent if every state is to be visited infinitely often.

A Markov chain is recurrent if every state is to be visited infinitely often.

Define the return time:

$$\tau_i = \inf\{n \geq 0 : X_n = i\}.$$

We say that the state is recurrent if

$$\mathbb{P}(\tau_i < \infty | X_1 = i) = 1.$$

If a state is not recurrent, it is transient.

We say that a state i is positively recurrent if

$$\mathbb{E}[\tau_i | X_1 = i] < \infty.$$

We say that a state i is positively recurrent if

$$\mathbb{E}[\tau_i | X_1 = i] < \infty.$$

If a recurrent state is not positive recurrent, it is null recurrent.

A probability mass function $p_\star$ is called M-invariant if

$$p_\star(i) = \sum_j M_{ij} p_\star(j).$$

A probability mass function $p_\star$ is called M-invariant if

$$p_\star(i) = \sum_j M_{ij} p_\star(j).$$

Equivalently

$$p_\star = p_\star M.$$

Theorem 1

If M is irreducible, then M has a unique invariant distribution if and only if it is positive recurrent.

Theorem 1

If M is irreducible, then M has a unique invariant distribution if and only if it is positive recurrent.

This is existence, we do not talk about convergence yet.

We will define reversibility through detailed balance condition.

We will define reversibility through detailed balance condition.

A Markov transition matrix M is reversible w.r.t. $p_\star$ if and only if for all $i, j \in \mathsf{X}$,

$$p_\star(i) M_{ij} = p_\star(j) M_{ji}.$$

We will define reversibility through detailed balance condition.

A Markov transition matrix M is reversible w.r.t. $p_\star$ if and only if for all $i, j \in \mathsf{X}$,

$$p_\star(i) M_{ij} = p_\star(j) M_{ji}.$$

This is called the detailed balance condition (we will discuss the continuous version)

Constructing a chain with stationary distribution $p_\star$ is ensured if detailed balance is satisfied since it implies $p_\star = p_\star M$.

We have seen how to construct chains with invariant distributions $p_\star$.

We have seen how to construct chains with invariant distributions $p_\star$.

However, the convergence of the chain $p_n \to p_\star$ requires one more ingredient: ergodicity.

We have seen how to construct chains with invariant distributions $p_\star$.

However, the convergence of the chain $p_n \to p_\star$ requires one more ingredient: ergodicity.

For this, we need a final ingredient: aperiodicity.

A state i is called aperiodic if

$$\{n > 0 : \mathbb{P}(X_{n+1} = i | X_1 = i) > 0\}$$

has no common divisor other than 1.

Definition 2

An irreducible Markov chain is called ergodic if it is positive recurrent and aperiodic.

Definition 2

An irreducible Markov chain is called ergodic if it is positive recurrent and aperiodic.

Ergodicity brings us the missing ingredient for the convergence: We can now ensure p_n to converge to $p_\star$.

If $(X_n)_{n\in\mathbb{N}}$ is an ergodic Markov chain with any initial p_0 and a Markov transition matrix M with $p_\star$ as its invariant distribution, then

$$\lim_{n\to\infty} p_n(i) = p_\star(i).$$

If $(X_n)_{n \in \mathbb{N}}$ is an ergodic Markov chain with any initial p_0 and a Markov transition matrix M with $p_\star$ as its invariant distribution, then

$$\lim_{n \to \infty} p_n(i) = p_\star(i).$$

Moreover, for $i, j \in \mathsf{X}$

$$\lim_{n \to \infty} \mathbb{P}(X_n = i | X_1 = j) = p_\star(i).$$

What about continuous state-space Markov chains, i.e., where X is uncountable, e.g., $\mathsf{X} = \mathbb{R}$?

What about continuous state-space Markov chains, i.e., where X is uncountable, e.g., $\mathsf{X} = \mathbb{R}$?

We will be mainly interested in the continuous case, however, the analogous concepts are defined in a much more complicated way.

What about continuous state-space Markov chains, i.e., where X is uncountable, e.g., $\mathsf{X} = \mathbb{R}$?

We will be mainly interested in the continuous case, however, the analogous concepts are defined in a much more complicated way.

We will not go into the details here, we will just now introduce the continuous state-space notation.

We assume now our state-space is uncountable, e.g., $\mathsf{X} = \mathbb{R}$.

What is a Markov chain?

The continuous case case

We assume now our state-space is uncountable, e.g., $\mathsf{X} = \mathbb{R}$.

We denote the initial *density* of the chain by $p_0(x)$.

What is a Markov chain?

The continuous case case

We assume now our state-space is uncountable, e.g., $\mathsf{X} = \mathbb{R}$.

We denote the initial *density* of the chain by $p_0(x)$.

The transition kernel is denoted $K(x_n|x_{n-1})$.

What is a Markov chain?

The continuous case case

We assume now our state-space is uncountable, e.g., $\mathsf{X} = \mathbb{R}$.

We denote the initial *density* of the chain by $p_0(x)$.

The transition kernel is denoted $K(x_n|x_{n-1})$.

The density of the chain at time n is denoted by $p_n(x_n)$.

A discrete-time Markov chain is a process $(X_n)_{n\in\mathbb{N}}$, when X is uncountable, satisfies:

What is a Markov chain?

The continuous case

A discrete-time Markov chain is a process $(X_n)_{n\in\mathbb{N}}$, when X is uncountable, satisfies:

$$p(x_n|x_{1:n-1}) = p(x_n|x_{n-1}) = K(x_n|x_{n-1}).$$

What is a Markov chain?

The continuous case

A discrete-time Markov chain is a process $(X_n)_{n \in \mathbb{N}}$, when X is uncountable, satisfies:

$$p(x_n | x_{1:n-1}) = p(x_n | x_{n-1}) = K(x_n | x_{n-1}).$$

$$X_0 \longrightarrow X_1 \longrightarrow X_2 \longrightarrow \cdots \longrightarrow X_n$$

We will again consider the time-homogeneous case, i.e. the transition kernel is time-independent.

We will again consider the time-homogeneous case, i.e. the transition kernel is time-independent. A Markov chain therefore can be defined entirely by its:

- Initial state (or initial distribution)
- Transition kernel

Imperial College London

The transition kernel is a density function $K(x_n|x_{n-1})$ for fixed x_{n-1}, i.e.,

$$\int_{\mathsf{X}} K(x_n|x_{n-1})\,\mathrm{d}x_n = 1.$$

Otherwise, it is a function of (x_n, x_{n-1}).

Consider the following Markov chain: $X_0 = 0$ and

$$K(x_n|x_{n-1}) = \mathcal{N}(x_n; ax_{n-1}, 1),$$

where $0 < a < 1$.

Consider the following Markov chain: $X_0 = 0$ and

$$K(x_n|x_{n-1}) = \mathcal{N}(x_n; ax_{n-1}, 1),$$

where $0 < a < 1$.
We can simulate this chain by:

$$\begin{aligned}
X_1 &\sim \mathcal{N}(0, 1) \\
X_2 &\sim \mathcal{N}(aX_1, 1) \\
X_3 &\sim \mathcal{N}(aX_2, 1) \\
&\vdots \\
X_n &\sim \mathcal{N}(aX_{n-1}, 1).
\end{aligned}$$

Simulation.

The Chapman-Kolmogorov equation for the continuous case

$$p(x_n|x_{n-k}) = \int_{\mathsf{X}} K(x_n|x_{n-1}) p(x_{n-1}|x_{n-k}) \, \mathrm{d}x_{n-1},$$

for $k > 1$.

Let $p_0(x)$ be the initial density such that $X_0 \sim p_0(x)$.

Then, the density of the chain at time n is given by

$$p_n(x_n) = \int_{\mathsf{X}} K(x_n|x_{n-1}) p_{n-1}(x_{n-1})\, \mathrm{d}x_{n-1}.$$

What is a Markov chain?

The continuous case: m-step transition kernel

It is useful for us to define the m-step transition kernel:

$$
\begin{aligned}
p(x_{m+n}|x_n) &= K^m(x_{m+n}|x_n), \\
&= \int_{\mathsf{X}} K(x_{m+n}|x_{m+n-1}) \cdots K(x_{n+1}|x_n)\, \mathrm{d}x_{m+n-1} \cdots \mathrm{d}x_{n+1}.
\end{aligned}
$$

What is a Markov chain?

Properties

We have the similar conditions of aperiodicity and irreducibility as in the discrete case, but,

- These are defined over *sets* rather than states.
- irreducibility is replaced by ϕ-irreducibility.
- aperiodicity is defined for sets

We have the similar conditions of aperiodicity and irreducibility as in the discrete case, but,

- These are defined over *sets* rather than states.
- irreducibility is replaced by ϕ-irreducibility.
- aperiodicity is defined for sets

We will not go into the details of these conditions for continuous space case.

A probability distribution $p_\star$ is called K-invariant if

$$p_\star(x) = \int_{\mathsf{X}} p_\star(x')K(x|x')\,\mathrm{d}x'.$$

Similar to the discrete case.

The detailed balance condition for the continuous case takes a similar form:

$$p_\star(x)K(x'|x) = p_\star(x')K(x|x').$$

The detailed balance condition for the continuous case takes a similar form:

$$p_\star(x)K(x'|x) = p_\star(x')K(x|x').$$

Note that this is a sufficient condition for stationarity of $p_\star$:

$$\begin{aligned}\int p_\star(x)K(x'|x)\mathrm{d}y &= \int p_\star(x')K(x|x')\mathrm{d}x', \\ \implies p_\star(x) &= \int K(x|x')p_\star(x')\mathrm{d}x',\end{aligned}$$

which implies $p_\star$ is K-invariant.

Consider the following Markov chain: $X_0 = 0$ and

$$K(x_n|x_{n-1}) = \mathcal{N}(x_n; ax_{n-1}, 1),$$

where $0 < a < 1$.

Consider the following Markov chain: $X_0 = 0$ and

$$K(x_n|x_{n-1}) = \mathcal{N}(x_n; ax_{n-1}, 1),$$

where $0 < a < 1$. Note that we can also write this as

$$X_n = aX_{n-1} + \epsilon_n,$$

where $\epsilon_n \sim \mathcal{N}(0, 1)$.

What is a Markov chain?

Example: Go back to Gaussian model

Prove that for

$$p_\star(x) = \mathcal{N}\left(x; 0, \frac{1}{1 - a^2}\right),$$

the detailed balance condition is satisfied for the kernel

$$K(x_n | x_{n-1}) = \mathcal{N}(x_n; a x_{n-1}, 1),$$

where $0 < a < 1$.

Prove that $K^m(x_{m+n}|x_n)$ is given by

$$K^m(x_{m+n}|x_n) = \mathcal{N}\left(x_{m+n}; a^m x_n, \frac{1-a^{2m}}{1-a^2}\right).$$

Then prove that

$$p_\star(x) = \lim_{m\to\infty} K^m(x|x'),$$

independent of x'.

See you tomorrow!