WARWICK
THE UNIVERSITY OF WARWICK

Lecture 5: Sample paths, ergodicity, and MCMC

Susana Gomes

October 21st, 2021

Plan for today

1. Sample paths
2. Some examples
3. Ergodicity
4. An application: MCMC

Sample paths (holding times)

Another useful way to characterise CTMCs is by looking at sample paths and their properties.

A **sample path** $t \mapsto X_t(\omega)$ is a function that describes the state of our CTMC at time t. It is a piecewise constant and right-continuous function by convention.

Before we actually see how to write down sample paths, we will look at another concept:
For $X_0 = x$, we define the **holding time** $W_x := \inf\{t > 0 : X_t \neq x\}$.

Proposition:

The holding time W_x is **exponentially distributed** with mean $\frac{1}{|g(x,x)|}$, i.e., $W_x \sim \mathrm{Exp}(|g(x,x)|)$.
If $|g(x,x)| > 0$, the chain jumps to $y \neq x$ after time W_x with probability $\frac{g(x,y)}{|g(x,x)|}$.

Proof

For any $t, u > 0$, we have

$$\mathbb{P}(W_x > t+u \mid W_x > t) = \mathbb{P}(W_x > t+u \mid X_t = x)$$

$W_x > t$ means $X_t = x$

$$= \mathbb{P}(W_x > u) \rightarrow \text{memoryless property}$$

$W_x > t$ means $X_s = x \ \forall\ 0 < s < t$. Since X_t is indep. of X_s for $0 < s < t$ we can say this.

$$\Rightarrow \mathbb{P}(W_x > t+u) = \mathbb{P}(W_x > t+u \mid W_x > t)\,\mathbb{P}(W_x > t)$$

Law of total prob.

$$= \mathbb{P}(W_x > u)\,\mathbb{P}(W_x > t)$$

same arg. as friday

$$\Rightarrow \mathbb{P}(W_x > t) = e^{\gamma t} \quad \text{with } \gamma = \frac{d}{dt}\mathbb{P}(W_x > t)\Big|_{t=0}$$

$$= \lim_{\Delta t \to 0} \frac{p(x,x) + o(t) + 1}{\Delta t} = q(x,x) \leq 0.$$

↑ same as friday

Proof (continued)

Condition on chain leaving x shortly:

$$\mathbb{P}(X_{t+\Delta t} = y \mid X_t = x, W_x < \Delta t) =$$

will explain this separately

$$= \frac{\mathbb{P}(X_{t+\Delta t} \mid X_t = x)}{\mathbb{P}(W_x < \Delta t \mid X_t = x)} =$$

$$= \lim_{\Delta t \to 0} \frac{p_{\Delta t}(x,y)}{1 - p_{\Delta t}(x,x)} = \lim_{\Delta t \to 0} \frac{\Delta t\, g(x,y)}{1 - (1 + \Delta t\, g(x,x))}$$

Friday

Friday

$$= \frac{g(x,y)}{-g(x,x)}$$

Sample paths (jump times)

Once we have the holding times defined, we can define **jump times**: $J_0, J_1, \ldots$. These are defined recursively as

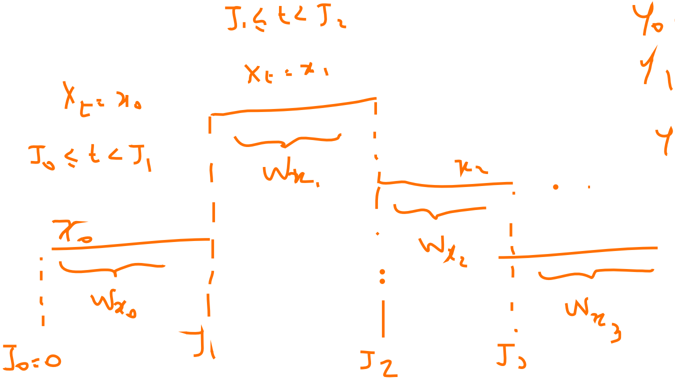
$$J_0 = 0 \quad \text{and} \quad J_{n+1} = \inf\{t > J_n : X_t \neq X_{J_n}\} \ .$$

- Jump times are an example of “**stopping times**" because we are working with right-continuous paths.
- This means that for all $t \geq 0$, the event $\{J_n \leq t\}$ depends **only** on $(X_s : 0 \leq s \leq t)$.
- This is justified by the **strong Markov property**: If we condition on a stopping time τ with $X_\tau = i$ then $X_{\tau+t}$ is independent of X_s for all $s \leq \tau$.
- i.e., subsequent holding times and jump probabilities are all independent.

Sample paths

Using the previous results, we can now define the **jump chain**:

$$(Y_n : n \in N_0) \quad \text{with} \quad Y_n := X_{J_n}$$



$Y_0 = x_0$

$Y_1 = X_{J_1} = x_1$

$Y_2 = X_{J_2} = x_2$

Sample paths

Using the previous results, we can now define the **jump chain**:

$$(Y_n : n \in N_0) \quad \text{with} \quad Y_n := X_{J_n}$$

This is a discrete-time Markov chain with transition matrix

$$p^Y(x,y) = \begin{cases} 0 & , \ x = y \\ g(x,y)/|g(x,x)| & , \ x \neq y \end{cases}$$

if $g(x,x) < 0$.

If $g(x,x) = 0$, we say (by convention) that

$$p^Y(x,y) = \delta_{x,y}.$$

Sample paths

A **sample path** is constructed by simulating the jump chain $(Y_n : n \in \mathbb{N}_0)$ together with independent **holding times** $(W_{Y_n} : n \in \mathbb{N}_0)$, so that

$$J_n = \sum_{k=0}^{n-1} W_{Y_k}.$$

A slide to sketch this if needed

Simulate CTMC X_t:

Start with X_0, define $Y_0 = X_0$

Compute $W_{Y_0} \to \cdot J_0 = W_{Y_0}$

use $g(Y_0, Y_0)$ and sample from $\exp(|g(Y_0, Y_0)|)$

Y_1 = next step of DTMC (advance using $p^Y(Y_0, y)$ from last slide)

$\to X_t = Y_1$ for $J_0 \le t < J_1$ (to be computed)

Compute W_{Y_1} (sample from $\exp(|g(Y_1, Y_1)|)$)

and $J_1 = W_{Y_0} + W_{Y_1}$

Compute Y_2, repeat.

Example 1 - Poisson Process

Suppose you want to model the arrival of customers to a waiting line. If we assume the following:

1. People arrive alone (never in groups).
2. The probability p that an arrival occurs during a time interval of (small) length Δt is proportional to Δt: $p = \lambda \Delta t$.
3. The number of arrivals on disjoint intervals is independent.

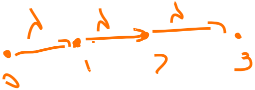
In this context, we would like to know, e.g., the law of the number of arrivals N_t in the interval $[0, t]$, or the number of arrivals per unit time.

The **Poisson process** is a good way to model this. Before we define it, we need to point out a couple of assumptions.

We can assume that people arrive in the waiting line completely at random. also, for **2.** to ve valid, we need to think in an “infinitesimal" sense, i.e., we should make sure that

$$\lim_{\Delta t \to 0} \frac{p}{\Delta t} = \lambda.$$

Poisson Process



A **Poisson process** with **rate** λ (short $\mathrm{PP}(\lambda)$) is a CTMC with

$$S = \mathbb{N}_0, \; X_0 = 0 \quad \text{and} \quad g(x, y) = \lambda\delta_{x+1,y} - \lambda\delta_{x,y}.$$

We can show that the $\mathrm{PP}(\lambda)$ has **stationary and independent increments** with

$$\mathbb{P}[X_{t+u} = n + k | X_u = n] = p_t(0, k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \quad \text{for all } u, t > 0, \; k, n \in \mathbb{N}_0.$$

This can be shown by dividing the time interval $[0, t]$ into intervals of length $\Delta t = t/n$ for n big enough and using properties of the Binomial law.

This is also related to the fact that $\pi_t(k) = p_t(0, k)$ solves the Master equation

$$\frac{d}{dt}\pi_t(k) = (\pi_t G)(k).$$

Example 2 - Birth-Death chains

A **birth-death chain** with **birth rates** α_x and **death rates** β_x is a CTMC with

$$S = \mathbb{N}_0 \quad \text{and} \quad g(x,y) = \alpha_x \delta_{x+1,y} + \beta_x \delta_{x-1,y} - (\alpha_x + \beta_x)\delta_{x,y},$$

where $\beta_0 = 0$.

These chains are used to model all sorts of things, from server queues to population sizes, to the evolution of an epidemic.

Special cases include

- **M/M/1 server queues**: $\alpha_x \equiv \alpha > 0$, $\beta_x \equiv \beta > 0$ for $x > 1$.
  e.g. a queue with Poisson arrivals but where one customer is served at a time, with random service time (following an exponential law).
- **M/M/∞ server queues**: $\alpha_x \equiv \alpha > 0$, $\beta_x = x\beta$.
  same as before but with immediate service.
- **population growth model**: $\alpha_x = x\alpha$, $\beta_x = x\beta$.

Ergodicity

A Markov process is called **ergodic** if it has a unique stationary distribution π and

$$p_t(x,y) = \mathbb{P}[X_t = y | X_0 = x] \to \pi(y) \quad \text{as } t \to \infty\,, \quad \text{for all } x, y \in S.$$

Theorem:

An **irreducible** (aperiodic) MC with finite state space is **ergodic**.

The proof of this follows from the Perron-Frobenius theorem:
We finished our lecture last Friday by saying that

$$p_t(x,y) = \sum_{i=1}^{|S|} \langle \delta_x | u_i \rangle \langle v_i | e^{\lambda_i t} \to \langle v_i | = \langle \pi | \quad \text{as } t \to \infty.$$

and this implies the theorem.
mention countably infinite state space.

Ergodicity

A very important result for ergodic Markov Chains is the Ergodic Theorem.

Theorem (Ergodic Theorem):

Consider an **ergodic Markov chain** with unique stationary distribution π. Then for every bounded function $f : S \to \mathbb{R}$ we have with probability 1

$$\frac{1}{T}\int_0^T f(X_t)\,dt \quad \text{or} \quad \frac{1}{N}\sum_{n=1}^{N} f(X_n) \ \to\ \mathbb{E}_\pi[f] \quad \text{as } T, N \to \infty\ .$$

For a proof of this theorem, you can check the book by Grimmett and Stirzaker (2001), chapter 9.5.

What this means is that stationary expectations can be approximated by time averages, which is the basis for **Markov chain Monte Carlo** (which we will see next).

An immediate example is that if we choose the indicator function $f = \mathbb{1}_x$ we get $\mathbb{E}_\pi[f] = \pi(x)$.

Markov Chain Monte Carlo (MCMC)

MCMC is used in several applications, when one needs to sample from some distribution π on a very large state space S. *(examples)*

Often, in these problems, one needs to compute complicated integrals which are not straightforward. Examples include:

- In general, compute **expectations**:

$$\mathbb{E}_\pi[f] = \sum_{x \in S} f(x)\pi(x) \quad \text{or} \quad \mathbb{E}_\pi[f] = \int f(x)\pi(x)\, dx$$

- In statistical mechanics, compute **Gibbs measures**:
  If $\pi(x) = \frac{1}{Z(\beta)} e^{-\beta H(x)}$, compute the **partition function**

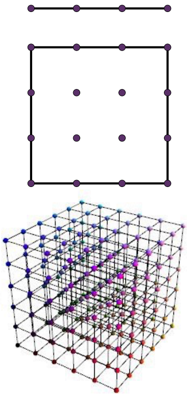
$$Z(\beta) = \sum_{x \in S} e^{-\beta H(x)}.$$

Why do we need Monte Carlo Methods?

Suppose you want to compute an integral in a domain $D = [0, 1]^d$ and want to use numerical quadrature to evaluate the integral.

1. Choose mesh of grid points within state space, with mesh-size h.
2. Evaluate $f(x_i)\pi(x_i)$ for every grid point x_i.
3. Use quadrature scheme to approximate integral.

- With the standard quadrature approach, error is typically $O(h^k)$, for some $k \geq 2$.
- The number of points evaluated is $M \sim O(h^{-d})$
- $\Rightarrow$ error $\sim O(M^{-k/d})$.



The computational cost grows exponentially with dimension (if we want to maintain the same error) This is usually known as curse of dimensionality.

MCMC

The reason that MCMC works is that we can use the **ergodic theorem** to estimate expectations by time averages! (together with reversibility)

For an MCMC algorithm, we need to:

- assume that $\pi(x) > 0$ for all $x \in S$ (otherwise restrict S).
- come up with a DTMC with transition function $p(x, y)$ (CTMC with generator $g(x, y)$) such that π is its stationary ditribution.
- This can be done, e.g., via **detailed balance**:

$$\pi(x)g(x,y) = \pi(y)g(y,x) \quad \text{(continuous)}$$

$$\pi(x)p(x,y) = \pi(y)p(y,x) \quad \text{(discrete)}.$$

- For example, for Gibbs measures, we know that

$$e^{-\beta H(x)}g(x,y) = e^{-\beta H(y)}g(y,x).$$

How does this actually work?

The main point of MCMC is that then we will *sample from the stationary distribution π.* To do this, we need to find the right $p(x, y)$. Suppose we have an associated distribution $q(\cdot|x)$ which is easy to sample.

1. Write $p(x, y) = q(x, y)\, a(x, y)$
2. Propose a move from x to y with probability $q(x, y)$
3. Accept this move from with probability $a(x, y)$.

To make sure this works, we need to check that the Markov chain we generated with $p(x, y)$ has π as its stationary distribution.

I will do this for the discrete case (and for one particular example), but the continuous case is analogous.

Metropolis-Hastings algorithm

This is one of the simplest MCMC algorithms you can think of.

Suppose that the chain is at the state X_n at time n. Then

1. Generate $Y \sim q(X_n, y)$.
2. Set $X_{n+1} = Y$ with probability $a(X_n, Y)$, where
$$a(x, y) = \min\left\{1, \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}\right\}$$
3. Otherwise, reject Y and set $X_{n+1} = X_n$..

We can easily prove that π is reversible with respect to the transition density of this DTMC, and therefore it is a stationary distribution!

We would then need to show it is ergodic for it all to work, but we won't do that here.

Space for notes if needed

$p(x, x) = ?$, $p(x,y) = ?$

1) $p(x,y) = q(x,y)\, a(x,y)$ ✓ (propose move from x to y <u>and</u> accept it)

2) $p(x,x) = q(x,x)\, a(x,x) + \sum_{z \in S} q(x,z)\,(1 - a(x,z))$

propose x & accept <u>or</u> propose $z \neq x \in S$ and reject it

$$\Rightarrow p(x,y) = q(x,y)\,a(x,y) + \delta_{x,y} \sum_{z \in S} q(x,z)\,(1 - a(x,z)).$$

→ Need to show it is reversible (to show stationary)

$$\pi(x)\, p(x,y) = \pi(x)\, q(x,y)\, a(x,y) = \pi(x) q(x,y) \min\left\{1, \frac{\pi(y) q(y,x)}{\pi(x) q(x,y)}\right\}$$

↑

$x=y$:!

$$= \min\left\{\pi(x)\, q(x,y),\ \pi(y)\, q(y,x)\right\} =$$

ow:

$$= \underbrace{\boxed{\min\left\{\frac{\pi(x)\, q(x,y)}{\pi(y)\, q(y,x)}, 1\right\}}}_{a(y,x)} \pi(y)\, q(y,x) = \pi(y)\, q(y,x)\, a(y,x)$$

$= \pi(y)\, p(y,x)$ so reversible!

Other examples

- **Independence sampler:** If $q(x, y) = q(y)$ (independent of the current state)

$$a(x, y) = \min\left\{1, \frac{\pi(y)q(x)}{\pi(x)q(y)}\right\}.$$