MATH-414 – Stochastic simulation

Lecture 10: Some theory of Markov Chains

Prof. Fabio Nobile

Outline

Convergence of discrete-state Markov Chains

Recall MCMC setting

- $\mathcal{X}$: state space
- π: target probability measure on $\mathcal{X}$, possibly known only up to a multiplicative constant (i.e. $\pi = C\tilde{\pi}$ and only $\tilde{\pi}$ is accessible).

Goals:

- sample from π
- Given $\psi : \mathcal{X} \to \mathbb{R}$ with finite first moment wrt π, compute $\mu = \mathbb{E}_\pi[\psi]$

Markov Chain Monte Carlo:

- Construct an ergodic Markov Chain $\{X_n,\ n \in \mathbb{N}_0\} \sim \mathrm{Markov}\,(\lambda, P)$ on $\mathcal{X}$ that has π as invariant distribution
- Approximate $\mu = \mathbb{E}_\pi[\psi]$ by ergodic estimator

$$\hat{\mu}_{N,b}^{MCMC} = \frac{1}{N}\sum_{i=1}^{N} \psi(X_{i+b})$$

Recall MCMC setting

In MCMC, the Markov chain $\{X_n\}_n$ is designed to have the prescribed invariant probability measure π

Relevant questions:

- Is the invariant probability measure unique?
- Does the ergodic estimator converge almost surely?

$$\frac{1}{N}\sum_{i=1}^{N}\psi(X_{i+b}) \overset{?}{\rightarrow} \mathbb{E}_\pi[\psi] \qquad a.s.$$

- Considering the n-th step distribution $\pi^{n,\lambda}(A) = \mathbb{P}_\lambda(X_n \in A)$, does $\pi^{n,\lambda}$ converge to π in some norm (or distance between probability measures)?
- In the theory of Markov Chains, another important question is the existence of an invariant probability measure. In MCMC this question is less relevant since an invariant probability meausre exists by construction.

Three key concepts: irreducibility, recurrence, aperiodicity

Notations

Consider a homogeneous Markov chain $\{X_n\}_n \sim \mathrm{Markov}\,(\lambda, P)$ on a discrete state space $\mathcal{X} = \{y_1, y_2, \ldots\}$

- $\lambda = (\lambda_1, \lambda_2, \ldots)$ is the initial measure, $\lambda_i = \mathbb{P}\,(X_0 = y_i)$
- P is the transition matrix, $P_{ij} = \mathbb{P}\,(X_{n+1} = y_j \mid X_n = y_i)$
- $P^{(n)}$ is the n-step transition matrix, $P^{(n)}_{ij} = \mathbb{P}\,(X_{m+n} = y_j \mid X_m = y_i)$ and $P^{(n)} = P^n$
- $\pi^{n,\lambda} = (\pi_1^{n,\lambda}, \pi_2^{n,\lambda}, \ldots)$ (row vector) is the probability mass function of X_n, $\pi_i^{n,\lambda} = \mathbb{P}_\lambda(X_n = y_i)$ and $\pi^{n,\lambda} = \pi^{n-1,\lambda} P = \lambda P^n$
- π is an invariant measure is $\pi = \pi P$

Let us define the following stopping times for a given set $A \subset \mathcal{X}$

- hitting time of A: $\tau_A = \inf\{n \geq 0 : \; X_n \in A\}$,
- return time to A: $\sigma_A = \inf\{n > 0 : \; X_n \in A\}$ (assuming $X_0 \in A$)
- successive return times to A: $\sigma_A^{(k)} = \inf\{n > \sigma_A^{(k-1)} : \; X_n \in A\}$, for $k \geq 1$

with the convention that $\tau_A = +\infty$ if $X_n \notin A$ for any n, $\sigma_A^{(0)} = 0$, $\sigma_A^{(1)} = \sigma_A$ and $\sigma_A^{(k)} = +\infty$ if $X_n \notin A$ for any $n > \sigma_A^{(k-1)}$.

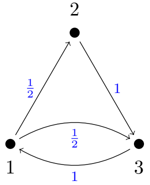
Irreducibility

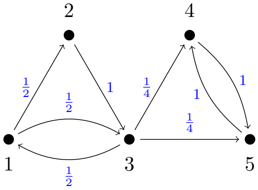
We say that y_i communicates with y_j if $\exists n : P_{ij}^{(n)} > 0$ (equivalently, $\mathbb{P}_i(X_n = y_j \text{ for some } n) > 0$)

Definition. *[**Irreducibility**] A Markov chain* $\mathrm{Markov}(\lambda, P)$ *is irreducible if every state communicates with every other state (i.e. for all* i, j*, there exists* $n > 0$ *such that* $P_{ij}^{(n)} > 0$*)*

We say that y_j is accessible if $\mathbb{P}_i(\sigma_j < \infty) > 0 \ \forall i$

Definition. *[**Irreducibility – equivalent definition**] A Markov chain* $\mathrm{Markov}(\lambda, P)$ *is irreducible if every state is accessible*





Recurrence and transience

Let $V_i = \sum_{n=0}^{\infty} \mathbb{1}_{\{X_n = y_i\}}$ be the number of visits to x_i.

We say that

- y_i is recurrent if $\mathbb{P}_i(V_i = \infty) = 1$
- y_i is transient if $\mathbb{P}_i(V_i = \infty) = 0$

Interesting facts: a state y_i is either recurrent or transient. Moreover

$$\begin{aligned}
\mathbb{P}_i(V_i = \infty) = 1 \quad &\iff \quad \mathbb{P}_i(\sigma_i < \infty) = 1 \quad \iff \quad \mathbb{E}_i[V_i] = \infty \\
\mathbb{P}_i(V_i = \infty) = 0 \quad &\iff \quad \mathbb{P}_i(\sigma_i < \infty) < 1 \quad \iff \quad \mathbb{E}_i[V_i] < \infty
\end{aligned}$$

Proof.

$$\begin{aligned}
\mathbb{P}_i(V_i > r+1) &= \mathbb{P}_i(\sigma_i^{(r+1)} < \infty) \\
&= \mathbb{P}_i(\sigma_i^{(r)} < \infty,\ \sigma_i^{(r+1)} - \sigma_i^{(r)} < +\infty) \\
&= \mathbb{P}_i(\sigma_i^{(r+1)} - \sigma_i^{(r)} < \infty \mid \sigma_i^{(r)} < \infty)\mathbb{P}_i(\sigma_i^{(r)} < \infty) \\
&= \mathbb{P}_i(\sigma_i^{(1)} < \infty)\mathbb{P}_i(\sigma_i^{(r)} < \infty) \text{ (by the strong Markov property)} \\
&= \mathbb{P}_i(\sigma_i < \infty)^{r+1} \xrightarrow{r \to \infty} \{0, 1\}
\end{aligned}$$

$\mathbb{P}_i(V_i = \infty) = \lim_{r\to\infty} \mathbb{P}_i(V_i > r) \in \{0, 1\}$

$\mathbb{E}_i[V_i] = \sum_r \mathbb{P}_i(V_i > r) \in \{C, \infty\}$

Recurrence and transience

Lemma

Let $\{X_n\} \sim \mathrm{Markov}(\lambda, P)$ *be irreducible. Then either all states are transient or recurrent.*

Proof.

Suppose y_i is transient and take $y_j \neq y_i$. Since P is irreducible, there exist $n, m > 0$: $P_{ij}^{(n)} > 0$ and $P_{ji}^{(m)} > 0$. Then for all $r \geq 0$,

$$P_{ii}^{(n+m+r)} = \sum_{\ell,k} P_{\ell i}^{(m)} P_{k\ell}^{(r)} P_{ik}^{(n)} \geq P_{ji}^{(m)} P_{jj}^{(r)} P_{ij}^{(n)}.$$

On the other hand, being y_i transient, we have $\mathbb{E}_i[V_i] < \infty$ and

$$\begin{aligned} \mathbb{E}_j[V_j] &= \mathbb{E}_j\Big[\sum_{r=0}^{\infty} \mathbb{1}_{\{X_r = y_j\}}\Big] = \sum_{r=0}^{\infty} \mathbb{P}_j(X_r = y_j) \\ &= \sum_{r=0}^{\infty} P_{jj}^{(r)} \leq \frac{1}{P_{ji}^{(m)} P_{ij}^{(n)}} \sum_{r=0}^{\infty} P_{ii}^{(m+n+r)} \leq \frac{1}{P_{ji}^{(m)} P_{ij}^{(n)}} \mathbb{E}_i[V_i] < \infty \end{aligned}$$

hence y_j is also transient.

Recurrence and transience

Definition. *An irreducible Markov chain* $\{X_n\} \sim \mathrm{Markov}(\lambda, P)$ *is recurrent if it has at least one recurrent state (equivalently if every state is recurrent)*

Exercise

- Show that the 1D random walk on $\mathbb{Z}$:

  $$\mathbb{P}(X_{n+1} = i+1 \mid X_n = i) = p, \quad \mathbb{P}(X_{n+1} = i-1 \mid X_n = i) = q = 1-p.$$

  is recurrent if $p = \frac{1}{2}$ and transient if $p \neq \frac{1}{2}$
- Show that the symmetric 2D random walk on $\mathbb{Z}^2$ (with $p = \frac{1}{4}$)

  $$\mathbb{P}(X_{n+1} = (i \pm 1, j) \mid X_n = (i,j)) = \mathbb{P}(X_{n+1} = (i, j \pm 1) \mid X_n = (i,j)) = \frac{1}{4}$$

  is recurrent
- Show that the symmetric random walk on $\mathbb{Z}^3$ ($p = \frac{1}{6}$) is transient.

Existence result

Theorem

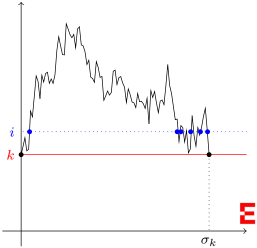
Let $\{X_n\}_n \sim \mathrm{Markov}\,(\lambda, P)$ *be irreducible and recurrent. Then* $\{X_n\}_n$ *has a unique invariant measure (not necessarily finite) up to a multiplicative constant.*

Constructive proof of existence: fix $y_k \in \mathcal{X}$ and define

$$\tilde{\pi}_i^k = \mathbb{E}_k\left[\sum_{n=0}^{\sigma_k - 1} \mathbb{1}_{\{X_n = x_i\}}\right], \qquad \tilde{\pi}_k^k = 1$$

(expected number of visits to y_i between two consecutive visits to y_k)

$\tilde{\pi}^k$ is an invariant measure and is unique up to a multiplicative factor.



Sketch of the proof (existence only)

Take $i \neq k$

$$\tilde{\pi}_i^k = \mathbb{E}_k\left[\sum_{n=0}^{\sigma_k - 1} \mathbb{1}_{\{X_n = y_i\}}\right] = \mathbb{E}_k\left[\sum_{n=1}^{\sigma_k} \mathbb{1}_{\{X_n = y_i\}}\right] = \mathbb{E}_k\left[\sum_{n=0}^{\sigma_k - 1} \mathbb{1}_{\{X_{n+1} = y_i\}}\right]$$

i.e. $\tilde{\pi}_i^k$ is invariant by a $+1$ right shift of the chain

$$\tilde{\pi}_i^k \overset{\text{by recurrence}}{=} \mathbb{E}_k\left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X_{n+1} = y_i,\, \sigma_k > n\}}\right] = \sum_{n=0}^{\infty} \mathbb{P}_k(X_{n+1} = y_i,\, \sigma_k > n)$$

$$= \sum_j \sum_{n=0}^{\infty} \underbrace{\mathbb{P}_k(X_{n+1} = y_i \mid X_n = y_j, \sigma_k > n)}_{= P_{ji} \text{ since } \{\sigma_k > n\} \text{ depends only on } X_0, \dots, X_n} \mathbb{P}_k(X_n = y_j, \sigma_k > n)$$

$$= \sum_j P_{ji} \underbrace{\sum_{n=0}^{\infty} \mathbb{P}_k(X_n = y_j, \sigma_k > n)}_{\tilde{\pi}_j^k} = \sum_j \tilde{\pi}_j^k P_{ji}$$

Positive recurrence

The measure $\tilde{\pi}^k$ is not necessarily finite. Normalizing constant:

$$\sum_i \tilde{\pi}_i^k = \sum_i \sum_{n=0}^{\infty} \mathbb{P}_k(X_n = y_i, \sigma_k > n) = \sum_{n=0}^{\infty} \mathbb{P}_k(\sigma_k > n) = \mathbb{E}_k[\sigma_k]$$

Definition. *An irreducible Markov chain $\{X_n\} \sim \mathrm{Markov}(\lambda, P)$ is positive recurrent (or simply positive) if $\mathbb{E}_i[\sigma_i] < \infty$ for at least one state $y_i \in \mathcal{X}$ and null recurrent otherwise.*

Theorem

Let $\{X_n\}_n \sim \mathrm{Markov}(\lambda, P)$ be irreducible, then $\{X_n\}_n$ has an invariant probability measure π if and only if it is positive recurrent. Moreover, in this case, π is unique and is given by

$$\pi_i = \frac{\tilde{\pi}_i^k}{\mathbb{E}_k[\sigma_k]} = \frac{1}{\mathbb{E}_i[\sigma_i]}.$$

(The last equality follows by simply taking $k = i$.)

By design, in MCMC the Markov chain has an invariant probability measure. Hence if it is irreducible, it is also positive recurrent and the invariant measure is unique.

Ergodicity

Theorem (Ergodic theorem)

Let $\{X_n\} \sim \mathrm{Markov}(\lambda, P)$ *be irreducible and positive recurrent, with invariant distribution* π. *Then, for any* π*-integrable function* $\psi : \mathcal{X} \to \mathbb{R}$

$$\mathbb{P}_\lambda \left(\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \psi(X_j) = \mathbb{E}_\pi[\psi] \right) = 1 \qquad \forall \lambda \in \mathcal{M}_1(\mathcal{X})$$

Idea of the proof.

Set $Y_r = \sum_{j=\sigma_k^{(r-1)}+1}^{\sigma_k^{(r)}} \psi(X_j)$. By strong Markov property, $\{Y_r\}_{r\geq 2}$ are iid!

Let $V_k(n) = \sum_{j=1}^{n-1} \mathbb{1}_{\{X_j = y_k\}}$ be the number of visits to y_k before n. Then

$$\frac{1}{V_k(n)-1} \sum_{r=2}^{V_k(n)} Y_r \xrightarrow{\text{a.s.}} \mathbb{E}\left[Y_r\right] \qquad \text{and} \qquad \mathbb{E}\left[Y_2\right] = \sum_i \psi(y_i)\tilde{\pi}_i^k = \frac{1}{\pi_k}\mathbb{E}_\pi[\psi]$$

$$\frac{1}{n}\sum_{j=1}^{n}\psi(X_j) = \underbrace{\frac{Y_1}{n}}_{\to 0} + \underbrace{\frac{V_k(n)-1}{n}}_{\to \pi_k} \underbrace{\frac{1}{V_k(n)-1}\sum_{r=2}^{V_k(n)} Y_r}_{\to \mathbb{E}[Y_r]} + \underbrace{\frac{1}{n}\sum_{j=\sigma_k^{V_k(n)}+1}^{n} \psi(X_j)}_{\to 0} \xrightarrow{\text{a.s.}} \pi_k \mathbb{E}\left[Y_r\right].$$

Central Limit Theorem

Exploiting the iid property of the random variables $\{Y_r\}_r$ one can also prove a CLT

Theorem (CLT for Markov Chains)

Let $\{X_n\} \sim \mathrm{Markov}\,(\lambda, P)$ *be irreducible, positive recurrent and with invariant distribution* π*. Let* $\psi : \mathcal{X} \to \mathbb{R}$ *be* π*-square integrable. Then*

$$\sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n}\psi(X_j) - \mathbb{E}_\pi[\psi]\right) \xrightarrow{d} N(0, C(\psi)).$$

with $C(\psi)$ *related to* $\mathbb{V}\mathrm{ar}[Y_r]$.

Convergence in distribution

We turn now to the question whether $\pi^{n,\lambda} \to \pi$ in some topology.

Example. Consider the (irreducible) transition matrix $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ having invariant distribution $\pi = (\frac{1}{2}, \frac{1}{2})$.

Take $\lambda = (1, 0)$. Then

$$\begin{aligned} \pi^{1,\lambda} &= \lambda P = (0, 1) \\ \pi^{2,\lambda} &= \pi^{1,\lambda} P = (1, 0) \\ &\dots \end{aligned}$$

The sequence $\{\pi^{n,\lambda}\}_n$ does not converge!

Yet the chain is irreducible and positive recurrent (since it has an invariant probability measure) and the ergodic theorem still applies!

Aperiodicity

Definition. *Given a transition matrix P, we say that a state y_i is aperiodic if $P_{ii}^{(n)} > 0$ for all sufficiently large n, or equivalently if the set $\{n > 0 : P_{ii}^n > 0\}$ has no common divisor other than 1.*

It P is irreducible and has an aperiodic state y_i, then all other states y_j are also aperiodic.

Indeed, if y_i is aperiodic, $\exists n_0 : \; P_{ii}^{(n)} > 0 \, \forall n \geq n_0$.
Take $j \neq i$. Since P is irreducible, $\exists n_1, n_2 : \; P_{ji}^{(n_1)} > 0, \; P_{ij}^{(n_2)} > 0$.
Hence, for any $m \geq m_0 = n_0 + n_1 + n_2$ we have

$$P_{jj}^{(m)} = \sum_{\ell,s} P_{j\ell}^{(n_1)} P_{\ell s}^{(m-n_1-n_2)} P_{sj}^{(n_2)} \geq P_{ji}^{(n_1)} P_{ii}^{(m-n_1-n_2)} P_{ij}^{(n_2)} > 0$$

Total variation

Let $(\mathcal{X}, \mathcal{B})$, be a measurable space and μ a (signed) measure on $(\mathcal{X}, \mathcal{B})$

Total Variation norm:

$$\|\mu\|_{\mathrm{TV}} = \sup_{A \in \mathcal{B}} \mu(A) - \inf_{A \in \mathcal{B}} \mu(A) = \sup_{\substack{f:\mathcal{X} \to \mathbb{R} \text{ meas.} \\ \|f\|_\infty \leq 1}} \int_{\mathcal{X}} f(x)\, \mu(dx)$$

If μ, ν are two probability measures on $(\mathcal{X}, \mathcal{B})$ then, it easy to see that

$$\|\mu - \nu\|_{\mathrm{TV}} = 2 \sup_{A \in \mathcal{B}} |\mu(A) - \nu(A)|$$

For a discrete state space $\mathcal{X} = \{y_1, y_2, \ldots\}$ and a signed measure $\mu = (\mu_1, \mu_2, \ldots)$, the TV norm reduces to the ℓ^1-norm

$$\|\mu\|_{\mathrm{TV}} = \sum_{y_i \in \mathcal{X}} |\mu_i| = \|\mu\|_{\ell^1}.$$

Convergence in Total Variation

Theorem

Let $\{X_n\}_n \sim \mathrm{Markov}\,(\lambda, P)$ be irreducible, aperiodic and positive recurrent with invariant distribution π. Let λ be any distribution on $\mathcal{X}$. Then, for $\pi_i^{n,\lambda} = \mathbb{P}_\lambda(X_n = y_i)$ it holds

$$\lim_{n\to\infty} \|\pi^{n,\lambda} - \pi\|_{TV} = \lim_{n\to\infty} \sum_i |\pi_i^{n,\lambda} - \pi_i| = 0.$$

Geometric / uniform ergodicity

Definition. *An irreducible, positive recurrent, aperiodic Markov chain $\{X_n\}_n$ with transition matrix P and invariant distribution π is*

- *geometrically ergodic if there exist $h : \mathcal{X} \to \mathbb{R}$, with $\mathbb{E}_\pi[h] < +\infty$ and $r \in (0, 1)$:*

$$\|\pi^{n,\delta_{x_i}} - \pi\|_{TV} \leq h(x_i) r^n \quad \textit{for all } x_i \in \mathcal{X},$$

- *uniformly ergodic if there exist $C > 0$ and $r \in (0, 1)$:*

$$\|\pi^{n,\delta_{x_i}} - \pi\|_{TV} \leq C r^n \quad \textit{for all } x_i \in \mathcal{X}.$$

Special case: finite state space, irreducible transition matric P

- $\{X_n\}_n$ is always recurrent and positive recurrent
- P has an eigenvalue $\lambda_1 = 1$ simple (Perrou-Frobenius theorem) and all other eigenvalues satisfy $|\lambda_i| < 1$, $i = 2, \dots, d$.
- $\{X_n\}_n$ is always uniformly ergodic and

$$\|\pi^{n,\delta_{x_i}} - \pi\|_{\mathrm{TV}} \leq C|\lambda_2|^n$$

with $|\lambda_2| = \max_{|\lambda_i|<1} |\lambda_i|$ if P is diagonalizable.
(If P is not diagonalizable, then $\|\pi^{n,\delta_{x_i}} - \pi\|_{\mathrm{TV}} \leq C(\epsilon)(|\lambda_2| + \epsilon)^n$ for $\epsilon > 0$ arbitrary small)

How the theory applies to Metropolis-Hastings

Algorithm: Metropolis-Hastings

```
Given: $\lambda$ (initial distribution), $Q$ (proposal), $\pi$ (target distribution)
Generate $X_0 \sim \lambda$
for $n = 0, 1, \ldots,$ do
    Generate candidate new state $\tilde{X}_{n+1} \sim Q_{X_n,:}$
    Generate $U \sim \mathcal{U}([0,1])$
    if $U \leq \alpha(X_n, \tilde{X}_{n+1})$ then
        set $X_{n+1} = \tilde{X}_{n+1}$    // $\tilde{X}_n$ accepted with prob. $\alpha(X_n, \tilde{X}_{n+1})$
    else
        set $X_{n+1} = X_n$    // $\tilde{X}_n$ rejected with prob. $1 - \alpha(X_n, \tilde{X}_{n+1})$
    end
end
```

Let $\alpha_i^* = \sum_j \alpha(i,j) Q_{ij}$. Transition matrix of the chain:

$$P_{ij} = \alpha(i,j) Q_{ij} + (1 - \alpha_i^*) \delta_{ij}.$$

How the theory applies to Metropolis-Hastings

- Irreducibility of P is implied by the irreducibility of Q. Hence we should always consider proposals Q that are irreducible.
- Positive recurrence is verified since P admits an inveriant probability measure by construction (remember that P has an invariant probability measure if and only if it is positive recurrent)
- Aperiodicity is satisfied automatically if $\alpha_i^* < 1$ for some i (positive probability of staying in state i in the next iteration – this brakes any periodicity).

  Only if $\alpha_i^* = 1$ for all i (e.g. in Gibb's sampler) we need to check that the chain is aperiodic