MATH-414 – Stochastic simulation

Lecture 11: Markov Chains on general state space Metropolis-Hastings algorithm

Prof. Fabio Nobile

EPFL MATHICSE

Outline

Markov Chains on a continuous state space

Metropolis-Hastings algorithm

Markov Chains on continuous state space

Let $\mathcal{X} \subset \mathbb{R}^d$ with Borel σ-algebra $\mathcal{B}(\mathcal{X})$ (we could even work on an arbitrary metric space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$)

Definition. *A Markov transition kernel on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is a function $P : \mathcal{X} \times \mathcal{B}(\mathcal{X}) \to [0, 1]$ s.t.*

1. *for all $x \in \mathcal{X}$, $P(x, \cdot)$ is a probability measure on $\mathcal{X}$,*
2. *for all $A \in \mathcal{B}(\mathcal{X})$, $P(\cdot, A)$ is measurable.*

The transition density associated to P, if it exists, is a function

$$p : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+, \qquad \int_{\mathcal{X}} p(x, y) dy = 1, \quad \forall x \in \mathcal{X}$$

such that $P(x, A) = \int_A p(x, y)\, dy$ for all $A \in \mathcal{B}(\mathcal{X})$.

Definition. *A sequence of random variables $\{X_n \in \mathcal{X}, n \geq 0\}$ is a homogeneous Markov chain with transition kernel P and initial distribution λ, in short $\{X_n\} \sim \mathrm{Markov}(\lambda, P)$ if*

- $X_0 \sim \lambda$
- $\mathbb{P}(X_{n+1} \in A \mid X_n, \ldots, X_0) = \mathbb{P}(X_{n+1} \in A \mid X_n) = P(X_n, A)$

Markov Chains on continuous state space

- n-step transition kernel

$$P^{(n)}(x, A) := \mathbb{P}(X_n \in A \mid X_0 = x) = \int_{\mathcal{X}} P^{(n-1)}(y, A) P(x, dy), \quad P^{(1)} = P$$

- n-step transition density

$$p^{(n)}(x, y) = \int_{\mathcal{X}} p^{(n-1)}(z, y) p(x, z)\, dz, \qquad p^{(1)} = p.$$

Markov transition operator

- Markov transition operator acting on measures (to the left): $\mathcal{P} : \mathcal{M}_1(\mathcal{X}) \to \mathcal{M}_1(\mathcal{X})$

$$\mu = \lambda\mathcal{P} \quad \Longrightarrow \quad \mu(A) = \int_{\mathcal{X}} P(x, A)\lambda(dx), \quad \forall A \in \mathcal{B}(\mathcal{X}).$$

implies

$$\lambda\mathcal{P}^2(A) = (\lambda\mathcal{P})\mathcal{P}(A) = \int_{\mathcal{X}} P(y, A)(\lambda\mathcal{P})(dy)$$
$$= \int_{\mathcal{X}} \int_{\mathcal{X}} P(y, A)P(x, dy)\lambda(dx) = \int_{\mathcal{X}} P^{(2)}(x, A)\lambda(dx)$$

and

$$\lambda\mathcal{P}^n(A) = \int_{\mathcal{X}} P^{(n)}(y, A)\lambda(dy)$$

- n-step distribution

$$\pi^{n,\lambda}(A) = \mathbb{P}_\lambda(X_n \in A) = \int_{\mathcal{X}} P^{(n)}(y, A)\lambda(dy) = \lambda\mathcal{P}^n(A)$$

Invariant measure and detailed balance

- A measure π is called invariant (or stationary) if

$$\pi = \pi\mathcal{P} = \int_{\mathcal{X}} P(y, \cdot)\pi(dy)$$

If f is the density of π and p the transition density, then

$$f(x) = \int_{\mathcal{X}} f(y)p(y, x)\, dy.$$

- A chain $\{X_n\}_{n=0} \sim \mathrm{Markov}\,(\lambda, P)$ is reversible if, for any $N > 0$, the chain $\{Y_n = X_{N-n}\}_{n=0}^{N} \sim \mathrm{Markov}\,(\lambda, P)$.
- (λ, P) are said to be in detailed balance if

$$\int_A P(x, B)\lambda(dx) = \int_B P(y, A)\lambda(dy), \quad \forall A, B \in \mathcal{B}(\mathcal{X}).$$

If ℓ denotes the density of λ, then $\ell(x)p(x, y) = \ell(y)p(y, x)$

- If (P, λ) are in detailed balance, then P is reversible and λ is an invariant measure. Indeed, $\forall B \in \mathcal{B}(\mathcal{X})$

$$(\lambda\mathcal{P})(B) = \int_{\mathcal{X}} P(x, B)\lambda(dx) = \int_B \underbrace{P(y, \mathcal{X})}_{=1}\lambda(dy) = \lambda(B).$$

Irreducibility

Definition. *A set $A \in \mathcal{B}(\mathcal{X})$ is accessible if $\mathbb{P}_x(\sigma_A < \infty) > 0$ for all $x \in \mathcal{X}$, where $\sigma_A = \inf\{n > 0 : X_n \in A\}$ is the return time to the set A.u*

Definition. *A Markov chain $\{X_n\}_n \sim \text{Markov}(\lambda, P)$ is irreducible if there exists a (σ-finite) measure φ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, called **irreducibility measure** such that any set $A \in \mathcal{B}(\mathcal{X})$, with $\varphi(A) > 0$ is accessible.*

The notion of irreducibility does not really depend on the measure φ as long as one irreducibilty measure exists.

Theorem

*If $\{X_n\}_n \sim \text{Markov}(\lambda, P)$ is irreducible for some irreducibility measure φ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, then there exists a probability measure ψ on $\mathcal{B}(\mathcal{X})$, called **maximal irreducibility measure** such that*

- *$\{X_n\}_n$ is ψ-irreducible*
- *For any other measure φ' on $\mathcal{B}(\mathcal{X})$ for which $\{X_n\}$ is φ'-irreducible, one has $\varphi' \gg \psi$ (i.e. for all $A \in \mathcal{B}(\mathcal{X})$, $\psi(A) = 0 \implies \varphi'(A) = 0$)*
- *Any invariant measure is a maximal irreducibility measure.*

Recurrence and a-periodicity

Definition. *A Markov chain $\{X_n\}_n \sim \mathrm{Markov}(\lambda, P)$ is aperiodic if for any $x \in \mathcal{X}$ and any accessible set $A \in \mathcal{B}(\mathcal{X})$*

$$\exists n_0 \geq 0 : \qquad P^{(n)}(x, A) > 0 \quad \forall n \geq n_0.$$

Let $A \in \mathcal{B}(\mathcal{X})$ and $V_A = \sum_{n \geq 0} \mathbb{1}_{\{X_n \in A\}}$ be the number of visits to A

Definition. *A Markov chain $\{X_n\}_n \sim \mathrm{Markov}(\lambda, P)$ is recurrent if it is irreducible and every accessible set A satisfies $\mathbb{E}_x[V_A] = \infty$, for all $x \in A$.*

It is Harris recurrent if it is irreducible and every accessible set A satisfies $\mathbb{P}_x(V_A = \infty) = 1$, for all $x \in A$;

Harris recurrence is stronger and implies recurrence. In the discrete state case, the two notions coincide.

Definition. *A Markov chain $\{X_n\}_n \sim \mathrm{Markov}(\lambda, P)$ is positive (recurrent) if it has an invariant probability measure.*

Theorem

An irreducible, recurrent Markov kernel P admits a non-zero invariant measure, unique up to a multiplicative constant.

If an irreducible Markov kernel P has an invariant probability measure (i.e. is positive), then it is recurrent.

Convergence of Markov chains

Theorem

Let $\{X_n\}_n$ be irreducible, poisitive, Harris recurrent, and aperiodic, with (unique) invariant distribution π. Then $\lim_{n\to\infty} \|\lambda \mathcal{P}^n - \pi\|_{TV} = 0$ *for any $\lambda \in \mathcal{M}_1(\mathcal{X})$.*

Theorem

Let $\{X_n\}_n$ be irreducible and positive, with invariant probability distribution π and let $\psi \in \mathcal{F}(\mathcal{X})$ be a π-integrable function with $\mathbb{E}_\pi[\psi] < \infty$. Then, for π-a.e. $x \in \mathcal{X}$

$$\mathbb{P}_x\left(\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \psi(X_j) = \mathbb{E}_\pi[\psi]\right) = 1.$$

Central limit theorems can be established as well.

Metropolis-Hastings algorithm on continuous state space

- State space $\mathcal{X} \subset \mathbb{R}^d$, target distribution π with density f.
- Take a proposal Markov transition kernel $Q : \mathcal{X} \times \mathcal{B}(\mathcal{X}) \to [0,1]$ with transition density $q : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ satisfying

$$Q(x, A) = \int_A q(x, y)dy, \qquad x \in \mathcal{X}, \;\; A \in \mathcal{B}(\mathcal{X})$$

  and $q(x, y) = 0 \;\Leftrightarrow q(y, x) = 0$.

- Define acceptance rate $\alpha : \mathcal{X} \times \mathcal{X} \to [0,1]$

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

  (if $q(x, y) = 0$ simply set $\alpha(x, y) = 0$)

- Given X_n
  - generate proposal state $\boldsymbol{Y}_{n+1} \sim Q(X_n, \cdot)$
  - with probabiliy $\alpha(X_n, Y_{n+1})$ accept the move and set $X_{n+1} = Y_{n+1}$. Otherwise, set $X_{n+1} = X_n$

Metropolis-Hastings algorithm

Algorithm: Metropolis-Hastings.

Given: λ (initial measure), q (proposal density), f (target density)

Generate $X_0 \sim \lambda$
for $n = 0, 1, \ldots,$ **do**
  Generate $Y_{n+1} \sim q(X_n, \cdot)$ `// proposal state`
  Generate $U \sim \mathcal{U}(0, 1)$
  if $U \leq \alpha(X_n, Y_{n+1})$ **then**
    set $X_{n+1} = Y_{n+1}$ `// accept proposal`
  else
    set $X_{n+1} = X_n$ `// reject proposal`
  end
end

Transition kernel of Metropolis-Hastings algorithm

- For $x \in \mathcal{X}$, overall acceptance probability of accepting the move being in x:

$$\alpha^*(x) = \int_{\mathcal{X}} \alpha(x, y) q(x, y)\, dy$$

- Transition density of Metropolis-Hastings algorithm:

$$p(x, y) = \alpha(x, y) q(x, y) + (1 - \alpha^*(x)) \delta_x(y), \quad x, y \in \mathcal{X},$$

where $\delta_x(y)$ is a Dirac mass in x.

- Markov transition kernel of Metropolis-Hastings algorithm:

$$P(x, A) = \int_A \alpha(x, y) q(x, y) dy + (1 - \alpha^*(x)) \mathbb{1}_A(x), \quad A \in \mathcal{B}(\mathcal{X}).$$

Detailed balance condition

Lemma

The transition kernel P of the Metropolis-Hastings algorithm is in detailed balance with the probability density f. Hence f is invariant probability density for P.

Proof Consider first the part of the kernel that has a density

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

Hence

$$\begin{aligned}\int_B P(x,A)f(x)\,dx &= \int_B \left(\int_A \alpha(x,y)q(x,y)\,dy\right) f(x)\,dx + \int_B (1-\alpha^*(x))\mathbb{1}_A(x)f(x)\,dx \\ &= \int_B\int_A f(y)\alpha(y,x)q(y,x)\,dy\,dx + \int_{A\cap B}(1-\alpha^*(x))f(x)\,dx \\ &= \int_A \left(\int_B \alpha(y,x)q(y,x)\,dx\right) f(y)\,dy + \int_A (1-\alpha^*(y))\mathbb{1}_B(y)f(y)\,dy \\ &= \int_A P(y,B)f(y)\,dy.\end{aligned}$$

On the convergence of the Metropolis-Hastings algorithm

- π-irreducibility. We have to check that each set $A \in \mathcal{B}(\mathcal{X})$ with $\pi(A) > 0$ is accessible.

  This will be guaranteed if the proposal density satisfies for instance $q(x, y) > 0$, $\forall x, y \in \mathcal{X}$

- Recurrence: this is guaranteed automatically by the fact that the chain has an invariant probability measure, π, hence it is positive. However, for convergence in total variation, we have to check Harris recurrence, which can be complicated.

- a-periodicity: This is guaranteed if $\alpha^*(x) < 1$ for any $x \in \mathcal{X}$. Indeed, in this case, by definition of accessible set,

$$\forall x \in \mathcal{X} \text{ and } A \in \mathcal{B}(\mathcal{X}) \text{ accessible}, \quad \exists n_0 : \; P^{(n_0)}(x, A) > 0.$$

  Hence

$$\begin{aligned} P^{(n_0+1)}(x, A) &= \int_{\mathcal{X}} P(y, A) P^{(n_0)}(x, dy) \geq \int_A P(y, A) P^{(n_0)}(x, dy) \\ &\geq \int_A (1 - \alpha^*(y)) P^{(n_0)}(x, dy) > 0. \end{aligned}$$

  Iterating the argument, we see that $P^{(n)}(x, A) > 0$ for any $n \geq n_0$