Markov Chain Monte Carlo
Theory and Practical applications
Chapters 2 and 3

Randal Douc and Sholom Schechtman

Télécom SudParis, Institut Polytechnique de Paris
randal.douc@telecom-sudparis.eu

Institut Mines-Télécom

Outline

1. Chap 2: Some recaps
2. Chap 2: Uniqueness of invariant probability measures
3. Chap 3: Dynamical systems
4. Chap 3:Markov chains and ergodicity

Outline

1. Chap 2: Some recaps
2. Chap 2: Uniqueness of invariant probability measures
3. Chap 3: Dynamical systems
4. Chap 3:Markov chains and ergodicity

Recall that a Markov kernel P is

1. π-invariant if $\pi P = \pi$
2. π-reversible if $\pi(\mathrm{d}x)P(x,\mathrm{d}y) = \pi(\mathrm{d}y)P(y,\mathrm{d}x)$
3. π-reversible implies π-invariance.

The Metropolis-Hastings algorithm

Input: n
Output: $X_0, \ldots, X_n$

- At $t = 0$, draw X_0 according to some arbitrary distribution
- For $t \leftarrow 0$ to $n-1$
  1. Draw independently $Y_{t+1} \sim$ Q $(X_t, \cdot)$ and $U_{t+1} \sim \mathrm{Unif}(0,1)$
  2. Set $X_{t+1} = \begin{cases} Y_{t+1} & \text{if } U_{t+1} \leqslant \alpha(X_t, Y_{t+1}) \\ X_t & \text{otherwise} \end{cases}$

where $\alpha(x,y) = \alpha^{MH}(x,y) = \min\left(\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1\right)$

Recall that a Markov kernel P is

1. π-invariant if $\pi P = \pi$
2. π-reversible if $\pi(\mathrm{d}x)P(x, \mathrm{d}y) = \pi(\mathrm{d}y)P(y, \mathrm{d}x)$
3. π-reversible implies π-invariance.

The Metropolis-Hastings algorithm

Input: n

Output: $X_0, \ldots, X_n$

- At $t = 0$, draw X_0 according to some arbitrary distribution
- For $t \leftarrow 0$ to $n - 1$
  1. Draw independently $Y_{t+1} \sim$ Q $(X_t, \cdot)$ and $U_{t+1} \sim \mathrm{Unif}(0, 1)$
  2. Set $X_{t+1} = \begin{cases} Y_{t+1} & \text{if } U_{t+1} \leqslant \alpha(X_t, Y_{t+1}) \\ X_t & \text{otherwise} \end{cases}$

where $\alpha(x, y) = \alpha^{MH}(x, y) = \min\left(\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1\right)$

Recall that a Markov kernel P is

1. π-invariant if $\pi P = \pi$
2. π-reversible if $\pi(\mathrm{d}x)P(x, \mathrm{d}y) = \pi(\mathrm{d}y)P(y, \mathrm{d}x)$
3. π-reversible implies π-invariance.

The Metropolis-Hastings algorithm

Input: n
Output: $X_0, \ldots, X_n$

- At $t = 0$, draw X_0 according to some arbitrary distribution
- For $t \leftarrow 0$ to $n - 1$
  1. Draw independently $Y_{t+1} \sim$ Q $(X_t, \cdot)$ and $U_{t+1} \sim \mathrm{Unif}(0, 1)$
  2. Set $X_{t+1} = \begin{cases} Y_{t+1} & \text{if } U_{t+1} \leqslant \alpha(X_t, Y_{t+1}) \\ X_t & \text{otherwise} \end{cases}$

where $\alpha(x, y) = \alpha^{MH}(x, y) = \min\left(\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1\right)$

Recall that a Markov kernel P is

1. π-invariant if $\pi P = \pi$
2. π-reversible if $\pi(\mathrm{d}x)P(x, \mathrm{d}y) = \pi(\mathrm{d}y)P(y, \mathrm{d}x)$
3. π-reversible implies π-invariance.

The Metropolis-Hastings algorithm

Input: n
Output: $X_0, \ldots, X_n$

- At $t = 0$, draw X_0 according to some arbitrary distribution
- For $t \leftarrow 0$ to $n - 1$
  1. Draw independently $Y_{t+1} \sim$ Q $(X_t, \cdot)$ and $U_{t+1} \sim \mathrm{Unif}(0, 1)$
  2. Set $X_{t+1} = \begin{cases} Y_{t+1} & \text{if } U_{t+1} \leqslant \alpha(X_t, Y_{t+1}) \\ X_t & \text{otherwise} \end{cases}$

where $\alpha(x, y) = \alpha^{MH}(x, y) = \min\left(\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1\right)$

Recall that a Markov kernel P is

1. π-invariant if $\pi P = \pi$
2. π-reversible if $\pi(\mathrm{d}x)P(x, \mathrm{d}y) = \pi(\mathrm{d}y)P(y, \mathrm{d}x)$
3. π-reversible implies π-invariance.

The Metropolis-Hastings algorithm

Input: n
Output: $X_0, \ldots, X_n$

- At $t = 0$, draw X_0 according to some arbitrary distribution
- For $t \leftarrow 0$ to $n - 1$
  1. Draw independently $Y_{t+1} \sim$ Q $(X_t, \cdot)$ and $U_{t+1} \sim \mathrm{Unif}(0, 1)$
  2. Set $X_{t+1} = \begin{cases} Y_{t+1} & \text{if } U_{t+1} \leqslant \alpha(X_t, Y_{t+1}) \\ X_t & \text{otherwise} \end{cases}$

where $\alpha(x, y) = \alpha^{MH}(x, y) = \min\left(\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1\right)$

Recall that a Markov kernel P is

1. π-invariant if $\pi P = \pi$
2. π-reversible if $\pi(\mathrm{d}x)P(x, \mathrm{d}y) = \pi(\mathrm{d}y)P(y, \mathrm{d}x)$
3. π-reversible implies π-invariance.

The Metropolis-Hastings algorithm

Input: n
Output: $X_0, \ldots, X_n$

- At $t = 0$, draw X_0 according to some arbitrary distribution
- For $t \leftarrow 0$ to $n - 1$
  1. Draw independently $Y_{t+1} \sim$ Q $(X_t, \cdot)$ and $U_{t+1} \sim \mathrm{Unif}(0, 1)$
  2. Set $X_{t+1} = \begin{cases} Y_{t+1} & \text{if } U_{t+1} \leqslant \alpha(X_t, Y_{t+1}) \\ X_t & \text{otherwise} \end{cases}$

  where $\alpha(x, y) = \alpha^{MH}(x, y) = \min\left(\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1\right)$

The Markov kernel associated to $\{X_n : n \in \mathbb{N}\}$ is given by

$$P^{MH}_{\langle \pi, Q \rangle}(x, \mathrm{d}y) = Q(x, \mathrm{d}y)\alpha(x, y) + \bar{\alpha}(x)\delta_x(\mathrm{d}y).$$

where $\bar{\alpha}(x) = 1 - \int_{\mathsf{X}} Q(x, \mathrm{d}y)\alpha(x, y)$.

Lemma

If the detailed balance condition

$$\pi(\mathrm{d}x)Q(x, \mathrm{d}y)\alpha(x, y) = \pi(\mathrm{d}y)Q(y, \mathrm{d}x)\alpha(y, x) \tag{1}$$

is satisfied, then $P^{MH}_{\langle \pi, Q \rangle}$ is π-reversible and hence, π-invariant.

- $\alpha^{MH}(x, y) = \min\left(\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1\right)$ or $\alpha^{b}(x, y) = \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)+\pi(y)q(y,x)}$ satisfy (1).
- For all α satisfying (1), we have $\alpha \leqslant \alpha^{MH}$. To be done on the blackboard.

The Markov kernel associated to $\{X_n : n \in \mathbb{N}\}$ is given by

$$P^{MH}_{\langle \pi, Q \rangle}(x, \mathrm{d}y) = Q(x, \mathrm{d}y)\alpha(x, y) + \bar{\alpha}(x)\delta_x(\mathrm{d}y).$$

where $\bar{\alpha}(x) = 1 - \int_{\mathsf{X}} Q(x, \mathrm{d}y)\alpha(x, y)$.

Lemma

If the detailed balance condition

$$\pi(\mathrm{d}x)Q(x, \mathrm{d}y)\alpha(x, y) = \pi(\mathrm{d}y)Q(y, \mathrm{d}x)\alpha(y, x) \quad (1)$$

is satisfied, then $P^{MH}_{\langle \pi, Q \rangle}$ is π-reversible and hence, π-invariant.

- $\alpha^{MH}(x, y) = \min\left(\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1\right)$ or $\alpha^{b}(x, y) = \frac{\pi(y)q(y,x)}{\pi(x)q(x,y) + \pi(y)q(y,x)}$ satisfy (1).
- For all α satisfying (1), we have $\alpha \leqslant \alpha^{MH}$. To be done on the blackboard.

The Markov kernel associated to $\{X_n : n \in \mathbb{N}\}$ is given by

$$P^{MH}_{\langle \pi, Q\rangle}(x, \mathrm{d}y) = Q(x, \mathrm{d}y)\alpha(x, y) + \bar{\alpha}(x)\delta_x(\mathrm{d}y).$$

where $\bar{\alpha}(x) = 1 - \int_{\mathsf{X}} Q(x, \mathrm{d}y)\alpha(x, y)$.

Lemma

If the detailed balance condition

$$\pi(\mathrm{d}x)Q(x, \mathrm{d}y)\alpha(x, y) = \pi(\mathrm{d}y)Q(y, \mathrm{d}x)\alpha(y, x) \qquad (1)$$

is satisfied, then $P^{MH}_{\langle \pi, Q\rangle}$ is π-reversible and hence, π-invariant.

- $\boxed{\alpha^{MH}(x, y) = \min\left(\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1\right)}$ or $\alpha^b(x, y) = \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)+\pi(y)q(y,x)}$ satisfy (1).
- For all α satisfying (1), we have $\alpha \leqslant \alpha^{MH}$. To be done on the blackboard.

The Markov kernel associated to $\{X_n : n \in \mathbb{N}\}$ is given by

$$P^{MH}_{\langle \pi, Q \rangle}(x, \mathrm{d}y) = Q(x, \mathrm{d}y)\alpha(x, y) + \bar{\alpha}(x)\delta_x(\mathrm{d}y).$$

where $\bar{\alpha}(x) = 1 - \int_{\mathsf{X}} Q(x, \mathrm{d}y)\alpha(x, y)$.

Lemma

If the detailed balance condition

$$\pi(\mathrm{d}x)Q(x, \mathrm{d}y)\alpha(x, y) = \pi(\mathrm{d}y)Q(y, \mathrm{d}x)\alpha(y, x) \tag{1}$$

is satisfied, then $P^{MH}_{\langle \pi, Q \rangle}$ is π-reversible and hence, π-invariant.

- $\boxed{\alpha^{MH}(x, y) = \min\left(\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1\right)}$ or $\alpha^b(x, y) = \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)+\pi(y)q(y,x)}$ satisfy (1).
- For all α satisfying (1), we have $\alpha \leqslant \alpha^{MH}$. To be done on the blackboard.

Outline

1. Chap 2: Some recaps
2. Chap 2: Uniqueness of invariant probability measures
3. Chap 3: Dynamical systems
4. Chap 3:Markov chains and ergodicity

Uniqueness under irreducibility assumptions

Proposition: Irreducible Markov kernels

Assume that there exists a non-null measure $\mu \in \mathsf{M}_+(\mathsf{X})$ satisfying the following property:

- ($\star$) For all $A \in \mathcal{X}$ such that $\mu(A) > 0$ and for all $x \in \mathsf{X}$, there exists $n \in \mathbb{N}$ such that $P^n(x, A) > 0$.

Then, P admits at most one invariant probability measure.

If condition ($\star$) is satisfied, we say that P is μ-irreducible.

Application: Metropolis-Hastings algorithms

Assume that

- $Q(x, \mathrm{d}y) = q(x, y)\lambda(\mathrm{d}y)$ and $\pi(\mathrm{d}x) = \pi(x)\lambda(\mathrm{d}x)$ with $q > 0$ and $\pi > 0$.

Then $P^{MH}_{\langle \pi, Q \rangle}$ admits π as its unique probability measure.

Uniqueness under irreducibility assumptions

Proposition: Irreducible Markov kernels

Assume that there exists a non-null measure $\mu \in \mathsf{M}_+(\mathsf{X})$ satisfying the following property:

- ($\star$) For all $A \in \mathcal{X}$ such that $\mu(A) > 0$ and for all $x \in \mathsf{X}$, there exists $n \in \mathbb{N}$ such that $P^n(x, A) > 0$.

Then, P admits at most one invariant probability measure.

If condition ($\star$) is satisfied, we say that P is μ-irreducible.

Application: Metropolis-Hastings algorithms

Assume that

- $Q(x, \mathrm{d}y) = q(x, y)\lambda(\mathrm{d}y)$ and $\pi(\mathrm{d}x) = \pi(x)\lambda(\mathrm{d}x)$ with $q > 0$ and $\pi > 0$.

Then $P^{MH}_{\langle \pi, Q \rangle}$ admits π as its unique probability measure.

Uniqueness under irreducibility assumptions

Proposition: Irreducible Markov kernels

Assume that there exists a non-null measure $\mu \in \mathsf{M}_+(\mathsf{X})$ satisfying the following property:

- ($\star$) For all $A \in \mathcal{X}$ such that $\mu(A) > 0$ and for all $x \in \mathsf{X}$, there exists $n \in \mathbb{N}$ such that $P^n(x, A) > 0$.

Then, P admits at most one invariant probability measure.

If condition $(\star)$ is satisfied, we say that P is μ-irreducible.

Application: Metropolis-Hastings algorithms

Assume that

- $Q(x, \mathrm{d}y) = q(x, y)\lambda(\mathrm{d}y)$ and $\pi(\mathrm{d}x) = \pi(x)\lambda(\mathrm{d}x)$ with $q > 0$ and $\pi > 0$.

Then $P^{MH}_{\langle \pi, Q \rangle}$ admits π as its unique probability measure.

Proof of the uniqueness of the invariant probability measure for irreducible Markov chains

The following lemma is useful for the proof...

Lemma

If P admits two distinct invariant probability measures, it also admits distinct invariant probability measures π_0 and π_1 that are mutually singular, i.e., such that there exists $A \in \mathcal{X}$ such that $\pi_0(A) = \pi_1(A^c) = 0$.

To be done on the blackboard.

Proof of the uniqueness of the invariant probability measure for irreducible Markov chains

The following lemma is useful for the proof...

Lemma

If P admits two distinct invariant probability measures, it also admits distinct invariant probability measures π_0 and π_1 that are mutually singular, i.e., such that there exists $A \in \mathcal{X}$ such that $\pi_0(A) = \pi_1(A^c) = 0$.

To be done on the blackboard.

Outline

1 Chap 2: Some recaps

2 Chap 2: Uniqueness of invariant probability measures

3 Chap 3: Dynamical systems

4 Chap 3:Markov chains and ergodicity

Dynamical systems

Definition

(Dynamical systems) A dynamical system $\mathcal{D}$ is a quadruplet $\mathcal{D} = (\Omega, \mathcal{F}, \mathbb{P}, T)$ where $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $T : \Omega \to \Omega$ is a measurable mapping such that $\boxed{\mathbb{P} = \mathbb{P} \circ T^{-1}}$.

Lemma

(Invariant sets) *The collection of sets* $\mathcal{I} = \{A \in \mathcal{F} \, : \, \mathbf{1}_A = \mathbf{1}_A \circ T\}$ *is a* σ*-field and any set in* $\mathcal{I}$ *is called an invariant set.*

Definition

(Ergodicity) A dynamical system $(\Omega, \mathcal{F}, \mathbb{P}, T)$ is said to be ergodic if invariant sets are $\mathbb{P}$-trivial that is if $A \in \mathcal{I}$ then either $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.

Dynamical systems

Definition

(Dynamical systems) A dynamical system $\mathcal{D}$ is a quadruplet $\mathcal{D} = (\Omega, \mathcal{F}, \mathbb{P}, T)$ where $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $T : \Omega \to \Omega$ is a measurable mapping such that $\boxed{\mathbb{P} = \mathbb{P} \circ T^{-1}}$.

Lemma

(Invariant sets) *The collection of sets $\mathcal{I} = \{A \in \mathcal{F} \, : \, \mathbf{1}_A = \mathbf{1}_A \circ T\}$ is a σ-field and any set in $\mathcal{I}$ is called an invariant set.*

Definition

(Ergodicity) A dynamical system $(\Omega, \mathcal{F}, \mathbb{P}, T)$ is said to be ergodic if invariant sets are $\mathbb{P}$-trivial that is if $A \in \mathcal{I}$ then either $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.

Dynamical systems

Definition

(Dynamical systems) A dynamical system $\mathcal{D}$ is a quadruplet $\mathcal{D} = (\Omega, \mathcal{F}, \mathbb{P}, T)$ where $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $T : \Omega \to \Omega$ is a measurable mapping such that $\boxed{\mathbb{P} = \mathbb{P} \circ T^{-1}}$.

Lemma

(Invariant sets) *The collection of sets* $\mathcal{I} = \{A \in \mathcal{F} \,:\, \mathbf{1}_A = \mathbf{1}_A \circ T\}$ *is a* σ*-field and any set in* $\mathcal{I}$ *is called an invariant set.*

Definition

(Ergodicity) A dynamical system $(\Omega, \mathcal{F}, \mathbb{P}, T)$ is said to be ergodic if invariant sets are $\mathbb{P}$-trivial that is if $A \in \mathcal{I}$ then either $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.

The Birkhoff theorem

Theorem

(The Birkhoff theorem) *Let* $\mathcal{D} = (\Omega, \mathcal{F}, \mathbb{P}, T)$ *be an* *ergodic dynamical system* *and let* $h \in \mathsf{L}_1(\Omega)$*. Then,*

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} h \circ T^k = \mathbb{E}[h] \;, \quad \mathbb{P} - a.s.$$

Outline

1. Chap 2: Some recaps
2. Chap 2: Uniqueness of invariant probability measures
3. Chap 3: Dynamical systems
4. Chap 3:Markov chains and ergodicity

Let S be the shift operator: if $\omega = (\omega_k)_{k\in\mathbb{N}} \in \mathsf{X}^{\mathbb{N}}$, we set $S(\omega) = \omega' \in \mathsf{X}^{\mathbb{N}}$ where $\omega'_k = \omega_{k+1}$ for all $k \in \mathbb{N}$.

Lemma (MC and dynamical systems)

Let P be a Markov kernel admitting an invariant probability measure π. Then, the quadruplet $(\mathsf{X}^{\mathbb{N}}, \mathcal{X}^{\otimes\mathbb{N}}, \mathbb{P}_\pi, S)$ is a dynamical system.

Theorem (MC and ergodicity)

Let P be a Markov kernel on $\mathsf{X} \times \mathcal{X}$. Assume that P admits a unique invariant probability measure π. Then, the dynamical system $(\mathsf{X}^{\mathbb{N}}, \mathcal{X}^{\otimes\mathbb{N}}, \mathbb{P}_\pi, S)$ is ergodic.

The proof of the Theorem will be done on the blackboard.

Let S be the shift operator: if $\omega = (\omega_k)_{k \in \mathbb{N}} \in \mathsf{X}^{\mathbb{N}}$, we set $S(\omega) = \omega' \in \mathsf{X}^{\mathbb{N}}$ where $\omega'_k = \omega_{k+1}$ for all $k \in \mathbb{N}$.

Lemma (MC and dynamical systems)

Let P be a Markov kernel admitting an invariant probability measure π. Then, the quadruplet $(\mathsf{X}^{\mathbb{N}}, \mathcal{X}^{\otimes \mathbb{N}}, \mathbb{P}_\pi, S)$ is a dynamical system.

Theorem (MC and ergodicity)

Let P be a Markov kernel on $\mathsf{X} \times \mathcal{X}$. Assume that P admits a unique invariant probability measure π. Then, the dynamical system $(\mathsf{X}^{\mathbb{N}}, \mathcal{X}^{\otimes \mathbb{N}}, \mathbb{P}_\pi, S)$ is ergodic.

The proof of the Theorem will be done on the blackboard.

Let S be the shift operator: if $\omega = (\omega_k)_{k \in \mathbb{N}} \in \mathsf{X}^{\mathbb{N}}$, we set $S(\omega) = \omega' \in \mathsf{X}^{\mathbb{N}}$ where $\omega'_k = \omega_{k+1}$ for all $k \in \mathbb{N}$.

Lemma (MC and dynamical systems)

Let P be a Markov kernel admitting an invariant probability measure π. Then, the quadruplet $(\mathsf{X}^{\mathbb{N}}, \mathcal{X}^{\otimes \mathbb{N}}, \mathbb{P}_\pi, S)$ is a dynamical system.

Theorem (MC and ergodicity)

Let P be a Markov kernel on $\mathsf{X} \times \mathcal{X}$. Assume that P admits a unique invariant probability measure π. Then, the dynamical system $(\mathsf{X}^{\mathbb{N}}, \mathcal{X}^{\otimes \mathbb{N}}, \mathbb{P}_\pi, S)$ is ergodic.

The proof of the Theorem will be done on the blackboard.

Theorem (The Birkhoff theorem for MC)

Let P be a Markov kernel admitting a unique invariant probability measure π. Then, for all $h \in \mathsf{F}(\mathsf{X}^{\mathbb{N}})$ such that $\mathbb{E}_\pi[|h|] < \infty$, we have

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} h(X_{k:\infty}) = \mathbb{E}_\pi[h]\,, \quad \mathbb{P}_\pi - a.s.$$

Corollary (LLN Starting from stationarity)

Let P be a Markov kernel admitting a unique invariant probability measure π. Then, for all $f \in \mathsf{F}(\mathsf{X})$ such that $\pi(|f|) = \int_{\mathsf{X}} \pi(\mathrm{d}x)|f(x)| < \infty$, we have

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} f(X_k) = \pi(f)\,, \quad \mathbb{P}_\pi - a.s. \tag{2}$$

Theorem (The Birkhoff theorem for MC)

Let P be a Markov kernel admitting a unique invariant probability measure π. Then, for all $h \in \mathsf{F}(\mathsf{X}^{\mathbb{N}})$ such that $\mathbb{E}_\pi[|h|] < \infty$, we have

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} h(X_{k:\infty}) = \mathbb{E}_\pi[h] \,, \quad \mathbb{P}_\pi - a.s.$$

Corollary (LLN Starting from stationarity)

Let P be a Markov kernel admitting a unique invariant probability measure π. Then, for all $f \in \mathsf{F}(\mathsf{X})$ such that $\pi(|f|) = \int_{\mathsf{X}} \pi(\mathrm{d}x)|f(x)| < \infty$, we have

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} f(X_k) = \pi(f) \,, \quad \mathbb{P}_\pi - a.s. \tag{2}$$

Corollary (Other starting points)

Let P be a Markov kernel admitting a unique invariant probability measure π. Then, for all $f \in \mathsf{F}(\mathsf{X})$ such that $\pi(|f|) = \int_{\mathsf{X}} \pi(\mathrm{d}x)|f(x)| < \infty$, we have for π-almost all $x \in \mathsf{X}$,

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} f(X_k) = \pi(f) \,, \quad \mathbb{P}_x - a.s. \tag{3}$$

Assume that $Q(x, \mathrm{d}y) = q(x,y)\lambda(\mathrm{d}y)$ and $\pi(\mathrm{d}y) = \pi(y)\lambda(\mathrm{d}y)$ where $q > 0$, $\pi > 0$ and λ is a σ-finite measure on $(\mathsf{X}, \mathcal{X})$.

Theorem

The Markov chain $\{X_n : n \in \mathbb{N}\}$ generated by the Metropolis-Hastings algorithm is such that: for all initial distributions $\nu \in \mathsf{M}_1(\mathsf{X})$ and all $f \in \mathsf{F}(\mathsf{X})$ such that $\pi(|f|) = \int_{\mathsf{X}} \pi(\mathrm{d}x)|f(x)| < \infty$,

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} f(X_k) = \pi(f)\,, \quad \mathbb{P}_\nu - a.s \tag{4}$$

What if P is not the Markov kernel of a Metropolis-Hastings algorithm?

Theorem

If P is a Markov kernel on $\mathsf{X} \times \mathcal{X}$ that admits a unique invariant probability measure π. Assume in addition that for all bounded functions h and all measures $\nu \in \mathsf{M}_1(\mathsf{X})$,

$$\lim_{n\to\infty} \nu P^n h = \pi(h) \tag{5}$$

Then, for all initial distributions $\nu \in \mathsf{M}_1(\mathsf{X})$ and all $f \in \mathsf{F}(\mathsf{X})$ such that $\pi(|f|) = \int_{\mathsf{X}} \pi(\mathrm{d}x)|f(x)| < \infty$,

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} f(X_k) = \pi(f) , \quad \mathbb{P}_\nu - a.s \tag{6}$$

What if P is not the Markov kernel of a Metropolis-Hastings algorithm?

Theorem

If P is a Markov kernel on $\mathsf{X} \times \mathcal{X}$ that admits a unique invariant probability measure π. Assume in addition that for all bounded functions h and all measures $\nu \in \mathsf{M}_1(\mathsf{X})$,

$$\lim_{n \to \infty} \nu P^n h = \pi(h) \tag{5}$$

Then, for all initial distributions $\nu \in \mathsf{M}_1(\mathsf{X})$ and all $f \in \mathsf{F}(\mathsf{X})$ such that $\pi(|f|) = \int_{\mathsf{X}} \pi(\mathrm{d}x)|f(x)| < \infty$,

$$\lim_{n \to \infty} n^{-1} \sum_{k=0}^{n-1} f(X_k) = \pi(f) , \quad \mathbb{P}_\nu - a.s \tag{6}$$