Introduction

Goal : For a given function f in some class of functions, approximate

$$\int \pi(\mathrm{d}x) f(x)$$

where the target distribution π is known up a multiplicative constant: $\pi(x) = C\tilde{\pi}(x)$ where $x \mapsto \tilde{\pi}(x)$ is known

- We use a Markov chain $(X_n)_{n\in\mathbb{N}}$ such that

$$\frac{1}{n}\sum_{i=0}^{n-1} f(X_i) \approx \int \pi(\mathrm{d}x) f(x) , \qquad n \text{ large} ,$$

- Theory of Markov chains: General definitions, invariant measures, ergodicity , Law of Large Numbers, geometric ergodicity, Central Limit theorems. 3 weeks.
- Practise of Markov chains: Metropolis-Hastings Markov chains and variants Pseudo marginal methods, Hamiltonian MCMC. Alternative methods (Sequential MC, Variational Inference, ABC). 3 weeks.

Introduction

Goal : For a given function f in some class of functions, approximate

$$\int \pi(\mathrm{d}x) f(x)$$

where the target distribution π is known up a multiplicative constant: $\pi(x) = C\tilde{\pi}(x)$ where $x \mapsto \tilde{\pi}(x)$ is known

- We use a Markov chain $(X_n)_{n\in\mathbb{N}}$ such that

$$\frac{1}{n}\sum_{i=0}^{n-1} f(X_i) \approx \int \pi(\mathrm{d}x) f(x) , \qquad n \text{ large} ,$$

- Theory of Markov chains: General definitions, invariant measures, ergodicity , Law of Large Numbers, geometric ergodicity, Central Limit theorems. 3 weeks.
- Practise of Markov chains: Metropolis-Hastings Markov chains and variants Pseudo marginal methods, Hamiltonian MCMC. Alternative methods (Sequential MC, Variational Inference, ABC). 3 weeks.

Introduction

Goal : For a given function f in some class of functions, approximate

$$\int \pi(\mathrm{d}x) f(x)$$

where the target distribution π is known up a multiplicative constant: $\pi(x) = C\tilde{\pi}(x)$ where $x \mapsto \tilde{\pi}(x)$ is known

- We use a Markov chain $(X_n)_{n\in\mathbb{N}}$ such that

$$\frac{1}{n}\sum_{i=0}^{n-1} f(X_i) \approx \int \pi(\mathrm{d}x) f(x) \,, \qquad n \text{ large} \,,$$

- Theory of Markov chains: General definitions, invariant measures, ergodicity , Law of Large Numbers, geometric ergodicity, Central Limit theorems. 3 weeks.
- Practise of Markov chains: Metropolis-Hastings Markov chains and variants Pseudo marginal methods, Hamiltonian MCMC. Alternative methods (Sequential MC, Variational Inference, ABC). 3 weeks.

Introduction

Goal : For a given function f in some class of functions, approximate

$$\int \pi(\mathrm{d}x) f(x)$$

where the target distribution π is known up a multiplicative constant: $\pi(x) = C\tilde{\pi}(x)$ where $x \mapsto \tilde{\pi}(x)$ is known

- We use a Markov chain $(X_n)_{n\in\mathbb{N}}$ such that

$$\frac{1}{n}\sum_{i=0}^{n-1} f(X_i) \approx \int \pi(\mathrm{d}x) f(x) , \qquad n \text{ large} ,$$

- Theory of Markov chains: General definitions, invariant measures, ergodicity , Law of Large Numbers, geometric ergodicity, Central Limit theorems. 3 weeks.
- Practise of Markov chains: Metropolis-Hastings Markov chains and variants Pseudo marginal methods, Hamiltonian MCMC. Alternative methods (Sequential MC, Variational Inference, ABC). 3 weeks.

Outline

Outline

1. Activities
2. Markov chains and Markov kernels
3. Finite dimensional laws
4. The canonical space
5. The Markov property

Outline

Definitions

Let $(\mathsf{X}, \mathcal{X})$ be a measurable space.

Definition (of a Markov kernel)

We say that $P : \mathsf{X} \times \mathcal{X} \to \mathbb{R}^+$ is a Markov kernel, if for all $(x, A) \in \mathsf{X} \times \mathcal{X}$,

- $y \mapsto P(y, A)$ is $\mathcal{X}/\mathcal{B}(\mathbb{R}^+)$ measurable,
- $B \mapsto P(x, B)$ is a probability measure on $(\mathsf{X}, \mathcal{X})$.

- In particular, $P(x, \mathsf{X}) = 1$ for all $x \in \mathsf{X}$.
- Recall if ν is a measure on $(\mathsf{X}, \mathcal{X})$, $A \mapsto \nu(A)$ is well-defined and we can define the integral associated to ν and we use the notation $\nu(f) = \int f(x)\nu(\mathrm{d}x)$,
- Since $P(x, \cdot)$ is a measure, we also use the infinitesimal notation: $P(x, \mathrm{d}y)$. For example,

$$P(x, A) = \int_{\mathsf{X}} \mathbf{1}_A(y) P(x, \mathrm{d}y) = \int_A P(x, \mathrm{d}y) .$$

Definitions

Let $(\mathsf{X}, \mathcal{X})$ be a measurable space.

Definition (of a Markov kernel)

We say that $P : \mathsf{X} \times \mathcal{X} \to \mathbb{R}^+$ is a Markov kernel, if for all $(x, A) \in \mathsf{X} \times \mathcal{X}$,

- $y \mapsto P(y, A)$ is $\mathcal{X}/\mathcal{B}(\mathbb{R}^+)$ measurable,
- $B \mapsto P(x, B)$ is a probability measure on $(\mathsf{X}, \mathcal{X})$.

- In particular, $P(x, \mathsf{X}) = 1$ for all $x \in \mathsf{X}$.
- Recall if ν is a measure on $(\mathsf{X}, \mathcal{X})$, $A \mapsto \nu(A)$ is well-defined and we can define the integral associated to ν and we use the notation $\nu(f) = \int f(x)\nu(\mathrm{d}x)$,
- Since $P(x, \cdot)$ is a measure, we also use the infinitesimal notation: $P(x, \mathrm{d}y)$. For example,

$$P(x, A) = \int_{\mathsf{X}} \mathbf{1}_A(y) P(x, \mathrm{d}y) = \int_A P(x, \mathrm{d}y) .$$

Definitions

Let $(\mathsf{X}, \mathcal{X})$ be a measurable space.

Definition (of a Markov kernel)

We say that $P : \mathsf{X} \times \mathcal{X} \to \mathbb{R}^+$ is a Markov kernel, if for all $(x, A) \in \mathsf{X} \times \mathcal{X}$,

- $y \mapsto P(y, A)$ is $\mathcal{X}/\mathcal{B}(\mathbb{R}^+)$ measurable,
- $B \mapsto P(x, B)$ is a probability measure on $(\mathsf{X}, \mathcal{X})$.

- In particular, $P(x, \mathsf{X}) = 1$ for all $x \in \mathsf{X}$.
- Recall if ν is a measure on $(\mathsf{X}, \mathcal{X})$, $A \mapsto \nu(A)$ is well-defined and we can define the integral associated to ν and we use the notation $\nu(f) = \int f(x)\nu(\mathrm{d}x)$,
- Since $P(x, \cdot)$ is a measure, we also use the infinitesimal notation: $P(x, \mathrm{d}y)$. For example,

$$P(x, A) = \int_{\mathsf{X}} \mathbf{1}_A(y) P(x, \mathrm{d}y) = \int_A P(x, \mathrm{d}y) .$$

Definitions

Let $(\mathsf{X}, \mathcal{X})$ be a measurable space.

Definition (of a Markov kernel)

We say that $P : \mathsf{X} \times \mathcal{X} \to \mathbb{R}^+$ is a Markov kernel, if for all $(x, A) \in \mathsf{X} \times \mathcal{X}$,

- $y \mapsto P(y, A)$ is $\mathcal{X}/\mathcal{B}(\mathbb{R}^+)$ measurable,
- $B \mapsto P(x, B)$ is a probability measure on $(\mathsf{X}, \mathcal{X})$.

- In particular, $P(x, \mathsf{X}) = 1$ for all $x \in \mathsf{X}$.
- Recall if ν is a measure on $(\mathsf{X}, \mathcal{X})$, $A \mapsto \nu(A)$ is well-defined and we can define the integral associated to ν and we use the notation $\nu(f) = \int f(x)\nu(\mathrm{d}x)$,
- Since $P(x, \cdot)$ is a measure, we also use the infinitesimal notation: $P(x, \mathrm{d}y)$. For example,

$$P(x, A) = \int_{\mathsf{X}} \mathbf{1}_A(y) P(x, \mathrm{d}y) = \int_A P(x, \mathrm{d}y) \; .$$

Let $\{X_k : k \in \mathbb{N}\}$ be a sequence of random variables on $(\Omega, \mathcal{G}, \mathbb{P})$ and taking values on X.

Definition (of a Markov chain)

We say that $\{X_k : k \in \mathbb{N}\}$ is a Markov chain with Markov kernel P and initial distribution $\nu \in \mathrm{M}_1(\mathsf{X})$ if and only if

1. for all $(k, A) \in \mathbb{N} \times \mathcal{X}$, $\mathbb{P}(X_{k+1} \in A | X_{0:k}) = P(X_k, A)$, $\mathbb{P}$-a.s.
2. $\mathbb{P}(X_0 \in A) = \nu(A)$.

Additional notation

Additional notation

For all $\mu \in \mathrm{M}_+(\mathsf{X})$, all Markov kernels P, Q on $\mathsf{X} \times \mathcal{X}$, and all measurable non-negative or bounded functions on h on X,

1. μP is the (positive) measure: $A \mapsto \mu P(A) = \int \mu(\mathrm{d}x) P(x, A)$,
2. PQ is the Markov kernel: $(x, A) \mapsto \int_{\mathsf{X}} P(x, \mathrm{d}y) Q(y, A)$,
3. Ph is the measurable function $x \mapsto \int_{\mathsf{X}} P(x, \mathrm{d}y) h(y)$.

- Example

$$\mu(P(Qh)) = (\mu P)(Qh) = (\mu(PQ))h = \mu((PQ)h)$$
$$= \int \cdots \int_{\mathsf{X}^3} \mu(\mathrm{d}x) P(x, \mathrm{d}y) Q(y, \mathrm{d}z) h(z) = \mu PQh$$

- Iterates of a kernel
  - define $P^0 = I$ where I is the identity kernel: $(x, A) \mapsto \mathbf{1}_A(x)$
  - set for $k \geqslant 0$, $P^{k+1} = P^k P$.

Additional notation

Additional notation

For all $\mu \in \mathrm{M}_+(\mathsf{X})$, all Markov kernels P, Q on $\mathsf{X} \times \mathcal{X}$, and all measurable non-negative or bounded functions on h on X,

❶ μP is the (positive) measure:
$A \mapsto \mu P(A) = \int \mu(\mathrm{d}x) P(x, A)$,

❷ PQ is the Markov kernel: $(x, A) \mapsto \int_{\mathsf{X}} P(x, \mathrm{d}y) Q(y, A)$,

❸ Ph is the measurable function $x \mapsto \int_{\mathsf{X}} P(x, \mathrm{d}y) h(y)$.

- Example

$$\mu(P(Qh)) = (\mu P)(Qh) = (\mu(PQ))h = \mu((PQ)h)$$
$$= \int \cdots \int_{\mathsf{X}^3} \mu(\mathrm{d}x) P(x, \mathrm{d}y) Q(y, \mathrm{d}z) h(z) = \mu PQh$$

- Iterates of a kernel
  - define $P^0 = I$ where I is the identity kernel: $(x, A) \mapsto \mathbf{1}_A(x)$
  - set for $k \geqslant 0$, $P^{k+1} = P^k P$.

Additional notation

Additional notation

For all $\mu \in \mathrm{M}_+(\mathsf{X})$, all Markov kernels P, Q on $\mathsf{X} \times \mathcal{X}$, and all measurable non-negative or bounded functions on h on X,

1. μP is the (positive) measure: $A \mapsto \mu P(A) = \int \mu(\mathrm{d}x) P(x, A)$,
2. PQ is the Markov kernel: $(x, A) \mapsto \int_{\mathsf{X}} P(x, \mathrm{d}y) Q(y, A)$,
3. Ph is the measurable function $x \mapsto \int_{\mathsf{X}} P(x, \mathrm{d}y) h(y)$.

- Example

$$\begin{aligned}\mu(P(Qh)) = (\mu P)(Qh) = (\mu(PQ))h = \mu((PQ)h) \\ = \int \cdots \int_{\mathsf{X}^3} \mu(\mathrm{d}x) P(x, \mathrm{d}y) Q(y, \mathrm{d}z) h(z) = \mu PQh\end{aligned}$$

- Iterates of a kernel
  - define $P^0 = I$ where I is the identity kernel: $(x, A) \mapsto \mathbf{1}_A(x)$
  - set for $k \geqslant 0$, $P^{k+1} = P^k P$.

Outline

Finite dimensional law

Let $\{X_k : k \in \mathbb{N}\}$ be a Markov chain with Markov kernel P and initial distribution $\nu \in \mathrm{M}_1(\mathsf{X})$

Lemma (The joint law)

For any $n \in \mathbb{N}$, the joint law of $X_{0:n}$ is

$$\nu(\mathrm{d}x_0) \prod_{i=1}^{n} P(x_{i-1}, \mathrm{d}x_i)$$

(with the convention that $\prod_{i=0}^{-1} = 1$). In particular, the law of X_n is νP^n.

Outline

1. let P be a Markov kernel on $\mathsf{X} \times \mathcal{X}$
2. let $\nu \in \mathrm{M}_1(\mathsf{X})$

Theorem

(The canonical space) *Given (1) and (2), there exists a unique probability measure $\mathbb{P}_\nu$ on the canonical space $(\mathsf{X}^{\mathbb{N}}, \mathcal{X}^{\otimes \mathbb{N}})$ such that*

- *under $\mathbb{P}_\nu$, the coordinate process $\{X_n : n \in \mathbb{N}\}$ is a Markov chain with Markov kernel P and initial distribution ν.*

1. We use the notation: $\mathbb{P}_x = \mathbb{P}_{\delta_x}$.
2. For any $A \in \mathcal{X}^{\otimes(n+1)}$

$$\mathbb{P}_\nu(X_{0:n} \in A) = \int_{\mathsf{X}} \nu(\mathrm{d}x_0)\mathbb{P}_{x_0}(X_{0:n} \in A).$$

3. We can replace n by ∞: for all $A \in \mathcal{X}^{\otimes \mathbb{N}}$,

$$\mathbb{P}_\nu(A) = \mathbb{P}_\nu(X_{0:\infty} \in A) = \int_{\mathsf{X}} \nu(\mathrm{d}x_0)\mathbb{P}_{x_0}(X_{0:\infty} \in A)$$
$$= \int_{\mathsf{X}} \nu(\mathrm{d}x_0)\mathbb{P}_{x_0}(A).$$

1. We use the notation: $\mathbb{P}_x = \mathbb{P}_{\delta_x}$.
2. For any $A \in \mathcal{X}^{\otimes(n+1)}$

$$\mathbb{P}_\nu(X_{0:n} \in A) = \int_{\mathsf{X}} \nu(\mathrm{d}x_0)\mathbb{P}_{x_0}(X_{0:n} \in A).$$

3. We can replace n by ∞: for all $A \in \mathcal{X}^{\otimes\mathbb{N}}$,

$$\mathbb{P}_\nu(A) = \mathbb{P}_\nu(X_{0:\infty} \in A) = \int_{\mathsf{X}} \nu(\mathrm{d}x_0)\mathbb{P}_{x_0}(X_{0:\infty} \in A) = \int_{\mathsf{X}} \nu(\mathrm{d}x_0)\mathbb{P}_{x_0}(A).$$

1. We use the notation: $\mathbb{P}_x = \mathbb{P}_{\delta_x}$.
2. For any $A \in \mathcal{X}^{\otimes(n+1)}$

$$\mathbb{P}_\nu(X_{0:n} \in A) = \int_{\mathsf{X}} \nu(\mathrm{d}x_0)\mathbb{P}_{x_0}(X_{0:n} \in A).$$

3. We can replace n by ∞: for all $A \in \mathcal{X}^{\otimes \mathbb{N}}$,

$$\begin{aligned}\mathbb{P}_\nu(A) = \mathbb{P}_\nu(X_{0:\infty} \in A) &= \int_{\mathsf{X}} \nu(\mathrm{d}x_0)\mathbb{P}_{x_0}(X_{0:\infty} \in A) \\ &= \int_{\mathsf{X}} \nu(\mathrm{d}x_0)\mathbb{P}_{x_0}(A).\end{aligned}$$

Outline

Theorem

(The Markov property) *For any $\nu \in \mathrm{M}_1(\mathsf{X})$, any non-negative or bounded function h on $\mathsf{X}^{\mathbb{N}}$ and any $k \in \mathbb{N}$,*

$$\mathbb{E}_\nu \left[h(X_{k:\infty}) | \mathcal{F}_k \right] = \mathbb{E}_{X_k}[h(X_{0:\infty})] \,, \quad \mathbb{P}_\nu - a.s. \tag{1}$$

where $\mathcal{F}_k = \sigma(X_{0:k})$.