Part 2: Monte Carlo and Markov chain Monte Carlo methods

Jean-Michel Marin

Université de Montpellier
Institut Montpelliérain Alexander Grothendieck (IMAG)

September, 22-27, Fréjus, France
1. Standard Monte Carlo method
2. Importance Sampling methods
3. Reminders and Additions on Markov Chains
4. Convergence of Markov chains
5. The Metropolis-Hastings algorithm
6. The Gibbs sampler
General definition: use of randomness to solve a problem centered on a calculation.
Standard Monte Carlo method

**General definition** use of randomness to solve a problem centered on a calculation

There is no consensus to give a more precise definition
General definition use of randomness to solve a problem centered on a calculation

There is no consensus to give a more precise definition

Methods that have been used for centuries: traces as far away as in Babylon and the Old Testament!
[1733, Buffon’s Needle] give an approximate value to $\pi$
[1733, Buffon’s Needle] give an approximate value to $\pi$

Throw a $l$ long needle on a floor of parallel slats that create $d$ widths with $l \leq d$
[1733, Buffon’s Needle] give an approximate value to $\pi$

Throw a $l$ long needle on a floor of parallel slats that create $d$ widths with $l \leq d$

If the needle is thrown uniformly on the ground (to be specified!), the probability that it intersects with one of the joins between the slats is $\frac{2l}{\pi d}$
Standard Monte Carlo method

[1733, Buffon’s Needle] give an approximate value to $\pi$

Throw a $l$ long needle on a floor of parallel slats that create $d$ widths with $l \leq d$

If the needle is thrown uniformly on the ground (to be specified!), the probability that it intersects with one of the joins between the slats is $\frac{2l}{\pi d}$

If you make several independent rolls and you note $p$ the proportion of tests that hit one of the straight lines forming the separations between the slats, $\pi$ can be estimated by $\frac{2l}{pd}$
Standard Monte Carlo method

[World War II, Los Alamos: Ulam, Metropolis and von Neumann] preparation of the first atomic bomb
Standard Monte Carlo method

[World War II, Los Alamos: Ulam, Metropolis and von Neumann] preparation of the first atomic bomb

The Monte Carlo appellation is due to Metropolis, inspired by Ulam’s interest in poker
Standard Monte Carlo method

[World War II, Los Alamos: Ulam, Metropolis and von Neumann] preparation of the first atomic bomb

The Monte Carlo appellation is due to Metropolis, inspired by Ulam’s interest in poker

Work at Los Alamos: directly simulate neutron dispersion and absorption problems for fissile materials
Standard Monte Carlo method

Theorem (strong law of large numbers) Let \((X_n)_{n \in \mathbb{N}}\) be an iid sequence of random variables with probability distribution \(f\) if \(\mathbb{E}_f(|X_i|) < \infty\), then

\[
\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \rightarrow_{ps} \mathbb{E}_f(X_1)
\]
Standard Monte Carlo method

**Theorem (central limit theorem)** Let $(X_n)_{n \in \mathbb{N}}$ be an iid sequence of random variables with probability distribution $f$

If $\mathbb{E}_f(|X_i|^2) < \infty$

$$\sqrt{n} \left( \frac{\bar{X}_n - \mathbb{E}_f(X_1)}{\sqrt{V_f(X_1)}} \right) \longrightarrow \mathcal{L} \ N(0, 1)$$
Standard Monte Carlo method

Target

$$\mathbb{E}_f(h(X)) = \int h(x) f(x) d\mu(x) < \infty$$

($f$ is the density of $X$ with respect to $\mu$)
Target

\[ \mathbb{E}_f(h(X)) = \int h(x)f(x)d\mu(x) < \infty \]

(f is the density of X with respect to \(\mu\))

**Standard Monte Carlo estimator of** \(\mathbb{E}_f(h(X))\)

\[ \frac{1}{n} \sum_{i=1}^{n} h(X_i) \]

where \(X_1, \ldots, X_n\) is an iid sample from \(f\)
Standard Monte Carlo method

Target

\[ \mathbb{E}_f(h(X)) = \int h(x)f(x)\,d\mu(x) < \infty \]

\(f\) is the density of \(X\) with respect to \(\mu\)

Standard Monte Carlo estimator of \(\mathbb{E}_f(h(X))\)

\[ \frac{1}{n} \sum_{i=1}^{n} h(X_i) \]

where \(X_1, \ldots, X_n\) is an iid sample from \(f\)
Standard Monte Carlo method

\[
\frac{1}{n} \sum_{i=1}^{n} h(X_i) \longrightarrow_{ps} \mathbb{E}_f(h(X))
\]
Standard Monte Carlo method

\[
\frac{1}{n} \sum_{i=1}^{n} h(X_i) \rightarrow_{ps} \mathbb{E}_f(h(X))
\]

\[
\mathbb{E}_{f \otimes n} \left( \frac{1}{n} \sum_{i=1}^{n} h(X_i) \right) = \mathbb{E}_f(h(X))
\]
Standard Monte Carlo method

\[ \mathbb{E}_f \otimes^n \left[ \frac{1}{n} \sum_{i=1}^{n} h(X_i) \right] = \frac{1}{n} \mathbb{E}_f(h(X)) \]
Standard Monte Carlo method

\[ \mathbb{V}_{f \otimes n} \left[ \frac{1}{n} \sum_{i=1}^{n} h(X_i) \right] = \frac{1}{n} \mathbb{V}_f(h(X)) \]

\[ \frac{1}{n} \left[ \frac{1}{n-1} \sum_{i=1}^{n} \left( h(X_i) - \frac{1}{n} \sum_{j=1}^{n} h(X_j) \right)^2 \right] \]

is an unbiased estimator of \( \mathbb{V}_f(h(X))/n \)
Standard Monte Carlo method

If \( \mathbb{E}_f(|h(X)|^2) < \infty \)

\[
\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} h(X_i) - \mathbb{E}_f(h(X)) \right) \rightarrow \mathcal{L} N(0, 1) 
\]
Standard Monte Carlo method

Convergence speed for various quadrature rules and for the Monte Carlo method in $s$ dimension and using $n$ points
Standard Monte Carlo method

Convergence speed for various quadrature rules and for the Monte Carlo method in $s$ dimension and using $n$ points

- Trapezoidal rule: $n^{-2/s}$
- Simpson rule: $n^{-4/s}$
- Gauss rule with $m$ points: $n^{-(2m-1)/s}$
- Monte-Carlo method: $n^{-1/2}$
Importance Sampling methods

Target

\[ E_f(h(X)) = \int h(x)f(x)d\mu(x) < \infty \]
Importance Sampling methods

Target

$$\mathbb{E}_f(h(X)) = \int h(x) f(x) d\mu(x) < \infty$$

We consider the probability density $g$ (with respect to $\mu$) such that: if $g(x) = 0$ then $f(x)|h(x)| = 0$
Importance Sampling methods

\[
\mathbb{E}_f(\mathbb{h}(X)) = \int h(x)f(x) \, d\mu(x) = \\
\int h(x) \frac{f(x)}{g(x)} g(x) \, d\mu(x) = \mathbb{E}_g \left[ h(X) \frac{f(X)}{g(X)} \right]
\]
Importance Sampling methods

\[ \mathbb{E}_f(h(X)) = \int h(x) f(x) \, d\mu(x) = \int h(x) \frac{f(x)}{g(x)} g(x) \, d\mu(x) = \mathbb{E}_g \left[ h(X) \frac{f(X)}{g(X)} \right] \]

Importance sampling estimator of \( \mathbb{E}_f(h(X)) \)

\[ \frac{1}{n} \sum_{i=1}^{n} h(X_i) \frac{f(X_i)}{g(X_i)} \]

where \( X_1, \ldots, X_n \) is an iid sample from \( g \)
Importance Sampling methods

If \( f|h| \) is absolutely continuous with respect to \( g \)

\[
\frac{1}{n} \sum_{i=1}^{n} h(X_i) \frac{f(X_i)}{g(X_i)} \rightarrow^{ps} \mathbb{E}_f(h(X))
\]

is convergent
Importance Sampling methods

**If** \( f|h| \) **is absolutely continuous with respect to** \( g \)

\[
\frac{1}{n} \sum_{i=1}^{n} h(X_i) \frac{f(X_i)}{g(X_i)} \xrightarrow{ps} E_f(h(X))
\]

**is convergent**

\[
\mathbb{E}_{g^\otimes n} \left( \frac{1}{n} \sum_{i=1}^{n} h(X_i) \frac{f(X_i)}{g(X_i)} \right) = E_f(h(X))
\]

**is unbiased**
Importance Sampling methods

\[
\mathbb{V}_{g^\otimes n} \left[ \frac{1}{n} \sum_{i=1}^{n} h(X_i) \frac{f(X_i)}{g(X_i)} \right] = \frac{1}{n} \mathbb{V}_{g} \left[ h(X) \frac{f(X)}{g(X)} \right]
\]

where

\[
\mathbb{V}_{g} \left[ h(X) \frac{f(X)}{g(X)} \right] = \mathbb{E}_{f} \left[ h(X)^2 \frac{f(X)}{g(X)} \right] - \left( \mathbb{E}_{f}(h(X)) \right)^2
\]
Importance Sampling methods

\[ \mathbb{V}_{g^\otimes n} \left[ \frac{1}{n} \sum_{i=1}^{n} h(X_i) \frac{f(X_i)}{g(X_i)} \right] = \frac{1}{n} \mathbb{V}_{g} \left[ h(X) \frac{f(X)}{g(X)} \right] \]

where

\[ \mathbb{V}_{g} \left[ h(X) \frac{f(X)}{g(X)} \right] = \mathbb{E}_{f} \left[ h(X)^2 \frac{f(X)}{g(X)} \right] - \left[ \mathbb{E}_{f}(h(X)) \right]^2 \]

is an unbiased estimator of \( \mathbb{V}_{g} \left[ h(X) \frac{f(X)}{g(X)} \right] / n \).
Importance Sampling methods

The importance function that minimise \( \mathbb{V}_g \left[ h(X) \frac{f(X)}{g(X)} \right] \) is

\[
g^*(x) = \frac{f(x)|h(x)|}{\int f(x)|h(x)| d\mu(x)}
\]

\( f|h| \) is absolutely continuous with respect to \( g^* \)
Importance Sampling methods

If $\mathbb{E}_g \left[ \left| h(X) \frac{f(X)}{g(X)} \right|^2 \right] = \mathbb{E}_f \left[ |h(X)|^2 \frac{f(X)}{g(X)} \right] < \infty$

$$\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} h(X_i) \frac{f(X_i)}{g(X_i)} - \mathbb{E}_f(h(X)) \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$
Importance Sampling methods

If \( \mathbb{E}_g \left[ \left| h(X) \frac{f(X)}{g(X)} \right|^2 \right] = \mathbb{E}_f \left[ |h(X)|^2 \frac{f(X)}{g(X)} \right] < \infty \)

\[
\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} h(X_i) \frac{f(X_i)}{g(X_i)} - \mathbb{E}_f(h(X)) \right) \rightarrow_{\mathcal{L}} \mathcal{N}(0, 1)
\]

\[
\frac{\sqrt{\mathbb{V}_g \left[ h(X)f(X)/g(X) \right]}}{\sqrt{V_f(h(X))/g(X)}}
\]

If \( f(x)/g(x) < M \) and \( \mathbb{V}_f(h(X)) < \infty \)

\[ \mathbb{E}_f \left[ |h(X)|^2 \frac{f(X)}{g(X)} \right] < \infty \]
Importance Sampling methods

There are many cases where the normalization constant of $f$ is unknown (Bayesian statistic)

$$f(x) = \frac{\tilde{f}(x)}{\int \tilde{f}(x) \, d\mu(x)} = \frac{\tilde{f}(x)}{c}$$
Importance Sampling methods

There are many cases where the normalization constant of $f$ is unknown (Bayesian statistic)

$$f(x) = \tilde{f}(x) / \int \tilde{f}(x) d\mu(x) = \tilde{f}(x)/c$$

Self-normalized importance sampling estimator of $\mathbb{E}_f(h(X))$

$$\frac{\sum_{i=1}^{n} h(X_i) \frac{f(X_i)}{g(X_i)}}{\sum_{i=1}^{n} \frac{f(X_i)}{g(X_i)}}$$

where $X_1, \ldots, X_n$ is an iid sample from $g$
Importance Sampling methods

If $f$ is absolutely continuous with respect to $g$,

$$
\frac{1}{n} \sum_{i=1}^{n} h(X_i) \frac{f(X_i)}{g(X_i)} \rightarrow_{ps} \mathbb{E}_f(h(X))
$$

is convergent
Importance Sampling methods

If $f$ is absolutely continuous with respect to $g$,

\[
\frac{1}{n} \sum_{i=1}^{n} h(X_i) \frac{f(X_i)}{g(X_i)} \Bigg/ \frac{1}{n} \sum_{i=1}^{n} \frac{f(X_i)}{g(X_i)} \to_{ps} E_f(h(X))
\]

is convergent

\[
\mathbb{E}_{g^\otimes n}\left( \frac{1}{n} \sum_{i=1}^{n} h(X_i) \frac{f(X_i)}{g(X_i)} \Bigg/ \frac{1}{n} \sum_{i=1}^{n} \frac{f(X_i)}{g(X_i)} \right) \neq \mathbb{E}_f(h(X))
\]
Importance Sampling methods

If \( \mathbb{E}_f \left[ |h(X)|^2 \frac{f(X)}{g(X)} \right] < \infty, \mathbb{E}_f \left[ \frac{f(X)}{g(X)} \right] < \infty, \)

\[
\sqrt{n} \left( \sum_{i=1}^{n} h(X_i) \frac{f(X_i)}{g(X_i)} \Bigg/ \sum_{i=1}^{n} \frac{f(X_i)}{g(X_i)} - \mathbb{E}_f(h(X)) \right) \longrightarrow \mathcal{L} \\
N \left( 0, \mathbb{E}_f \left[ [h(X) - \mathbb{E}_f(h(X))]^2 \frac{f(X)}{g(X)} \right] \right)
\]
Importance Sampling methods

If \( \mathbb{E}_f \left[ |h(X)|^2 \frac{f(X)}{g(X)} \right] < \infty, \mathbb{E}_f \left[ \frac{f(X)}{g(X)} \right] < \infty, \)

\[
\sqrt{n} \left( \sum_{i=1}^{n} h(X_i) \frac{f(X_i)}{g(X_i)} \right) / \sum_{i=1}^{n} \frac{f(X_i)}{g(X_i)} - \mathbb{E}_f (h(X)) \longrightarrow \mathcal{L} \]

\[
N \left( 0, \mathbb{E}_f \left( [h(X) - \mathbb{E}_f (h(X))]^2 f(X) / g(X) \right) \right)
\]

The importance function that minimise
\( \mathbb{E}_f \left( [h(X) - \mathbb{E}_f (h(X))]^2 f(X) / g(X) \right) \) is

\[
g^\#(x) = \frac{f(x)|h(x) - \mathbb{E}_f (h(X))|}{\int f(x)|h(x) - \mathbb{E}_f (h(X))|d\mu(x)}.
\]
Definition
A Markov chain is a random process \((X_k)_{k \in \mathbb{N}}\) such that

\[
P(X_k \in A | X_0 = x_0, \ldots, X_{k-1} = x_{k-1}) = P(X_k \in A | X_{k-1} = x_{k-1})
\]
Definition
A Markov chain is a random process \((X_k)_{k \in \mathbb{N}}\) such that

\[
P(X_k \in A | X_0 = x_0, \ldots, X_{k-1} = x_{k-1}) =
\]

\[
P(X_k \in A | X_{k-1} = x_{k-1})
\]

The Markov chain is homogenous if \(P(X_k \in A | X_{k-1} = x)\) does not depend on \(k\).
Reminders and Additions on Markov Chains

Example: random walk

\((X_k)_{k \in \mathbb{N}}\) such that

\[ X_0 \sim \nu \]

and

\[ X_k = X_{k-1} + \varepsilon_k, \quad \forall k \in \mathbb{N}^* \]

where \(\varepsilon_1, \ldots\) is a random process with iid variables and probability distribution \(\mathcal{L}\)
Reminders and Additions on Markov Chains

**Definition** A (transition) kernel on \((\Omega, \mathcal{A})\) is an application \(P: (\Omega, \mathcal{A}) \rightarrow [0, 1]\) such that

1) \(\forall A \in \mathcal{A}, P(\cdot, A)\) is measurable
2) \(\forall x \in \Omega, P(x, \cdot)\) is a probability distribution on \((\Omega, \mathcal{A})\)
Reminders and Additions on Markov Chains

**Definition** A (transition) kernel on \((\Omega, \mathcal{A})\) is an application \(P : (\Omega, \mathcal{A}) \rightarrow [0, 1]\) such that

1) \(\forall A \in \mathcal{A}, P(\cdot, A)\) is measurable

2) \(\forall x \in \Omega, P(x, \cdot)\) is a probability distribution on \((\Omega, \mathcal{A})\)

\((X_k)_{k \in \mathbb{N}}\) is an homogenous Markov chain with kernel \(P\) if

\[P(X_k \in A | X_{k-1} = x) = P(x, A), \quad \forall x \in \Omega, \quad \forall A \in \mathcal{A}.\]

Reminders and Additions on Markov Chains

For the random walk if $\mathcal{L} = \mathcal{N}(0, \sigma^2)$, $(X_k)_{k \in \mathbb{N}}$ is an homogeneous Markov chain with kernel

$$P(x, A) = \int_A \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y - x)^2\right) \, dy$$
Let \( (X_k)_{k \in \mathbb{N}} \) be an homogenous Markov chain with kernel \( P \) and initial distribution \( X_0 \sim \nu \), we note

- \( P_\nu \) the distribution of the chain \( (X_k)_{k \in \mathbb{N}} \)
- \( \nu P^k \) the distribution of \( X_k : \forall A \in \mathcal{A}, \nu P^k(A) = \mathbb{P}(X_k \in A) \)
- \( P^k(x, A) = \mathbb{P}(X_k \in A | X_0 = x) \)
Reminders and Additions on Markov Chains

Let $\Pi$ be a probability distribution on $(\Omega, \mathcal{A})$
Reminders and Additions on Markov Chains

Let $\Pi$ be a probability distribution on $(\Omega, \mathcal{A})$

We can simulate $\Pi$ in an approximate way using a homogeneous Markov chain
Reminders and Additions on Markov Chains

Let $\Pi$ be a probability distribution on $(\Omega, \mathcal{A})$

We can simulate $\Pi$ in an approximate way using a homogeneous Markov chain

To do this, one must be able to build a $P$ kernel such that for any initial $\nu$, $\nu P^k \rightarrow_{VT} \Pi$
Reminders and Additions on Markov Chains

Let \( \Pi \) be a probability distribution on \( (\Omega, \mathcal{A}) \)

We can simulate \( \Pi \) in an approximate way using a homogeneous Markov chain

To do this, one must be able to build a \( \mathcal{P} \) kernel such that for any initial \( \nu \), \( \nu \mathcal{P}^k \xrightarrow{\text{VT}} \Pi \)

Total variation convergence

\[
\|\nu \mathcal{P}^k - \pi\|_{\text{VT}} = \sup_{A \in \mathcal{A}} |\nu \mathcal{P}^k(A) - \pi(A)|
\]
Reminders and Additions on Markov Chains

Let \( \Pi \) be a probability distribution on \((\Omega, \mathcal{A})\)

We can simulate \( \Pi \) in an approximate way using a homogeneous Markov chain

To do this, one must be able to build a \( P \) kernel such that for any initial \( \nu \), \( \nu P^k \rightarrow_{VT} \Pi \)

Total variation convergence

\[
||\nu P^k - \pi||_{VT} = \sup_{A \in \mathcal{A}} |\nu P^k(A) - \pi(A)|
\]

Typically

\[
\lim_{k \rightarrow \infty} \nu P^k(A) = \pi(A)
\]
Reminders and Additions on Markov Chains

Definition

- $P$ is $\Pi$-irreducible if $\forall x \in \Omega$ and $\forall A \in \mathcal{A}$ such that $\Pi(A) > 0$, $\exists k(= k(x, A))$ tel que $P^k(x, A) > 0$

- $P$ is $\Pi$-invariant iff $\Pi P = \Pi$

$$\Pi P(A) = \int \Pi(dx_0)P(x_0, A) = \int_A \Pi(dx)$$

- $P$ is $\Pi$-reversible iff $\forall A, B \in \mathcal{A}$,

$$\int_A P(x, B)\Pi(dx) = \int_B P(x, A)\Pi(dx)$$
If $P$ est $\Pi$-reversible then $P$ is $\Pi$-invariant
Reminders and Additions on Markov Chains

If $P$ est $\Pi$-reversible then $P$ is $\Pi$-invariant

Indeed if $P$ is $\Pi$-reversible, $\forall B \in \mathcal{A}$,

$$
\int_\Omega P(x, B) \Pi(dx) = \int_B P(x, \Omega) \Pi(dx) = \int_B \Pi(dx)
$$
**Reminders and Additions on Markov Chains**

**Definition**

- $P$ is periodic with period $d \geq 2$ if there exists a partition $\Omega_1, \ldots, \Omega_d$ of $\Omega$ such that $\forall x \in \Omega_i$, $P(x, \Omega_{i+1}) = 1$, $\forall i$ with the convention $d + 1 = 1$

- A chain $\Pi$-irreducible and $\Pi$-invariant is recurrent if $\forall A \in \mathcal{A}$ such that $\pi(A) > 0$
  
  1) $\forall x \in \Omega$, $\mathbb{P}(X_k \in A \text{ infinitely often} | X_0 = x) > 0$
  2) $\exists x \in \Omega$, $\mathbb{P}(X_k \in A \text{ infinitely often} | X_0 = x) = 1$

- The chain is Harris-recurrent if 2) is verified for all $x \in \Omega$

- The chain is ergodic if it is Harris-recurrent and aperiodic
If $P$ is $\Pi$-irreducible and $\Pi$-invariant then $P$ is recurrent
Convergence of Markov chains

If $P$ is $\Pi$-irreducible and $\Pi$-invariant then $P$ is recurrent

In that case, the invariant measure is unique (up to a multiplicative constant)
Convergence of Markov chains

If $P$ is $\Pi$-irreducible and $\Pi$-invariant then $P$ is recurrent

In that case, the invariant measure is unique (up to a multiplicative constant)

The chain is said to be positive recurrent if the invariant measure is a probability distribution
**Theorem** Suppose that $P$ is $\Pi$-irreducible et $\Pi$-invariant, then $P$ is positive recurrent and $\Pi$ is the unique invariant distribution of $P$. If $P$ est Harris-recurrent et aperiodic (ergodic) then

$$\nu P^k \longrightarrow_{VT} \Pi$$
Theorem Suppose that $P$ is $\Pi$-irreducible et $\Pi$-invariant, then $P$ is positive recurrent and $\Pi$ is the unique invariant distribution of $P$. If $P$ est Harris-recurrent et aperiodic (ergodic) then

$$\nu P^k \rightarrow_{VT} \Pi$$

The Harris-recurrence condition is difficult to obtain
Theorem Suppose that $P$ is $\Pi$-irreducible et $\Pi$-invariant, then $P$ is positive recurrent and $\Pi$ is the unique invariant distribution of $P$. If $P$ est Harris-recurrent et aperiodic (ergodic) then

$$\nu P^k \rightarrow_{VT} \Pi$$

The Harris-recurrence condition is difficult to obtain

It is satisfied for two main families of simulators: the Gibbs sampler and the Metropolis-Hastings algorithm
Theorem If the Markov chain $(X_k)_{k \in \mathbb{N}}$ is ergodic with stationary distribution $\Pi$ and if $h$ is a real function such that $\mathbb{E}_\Pi(|h(X)|) < \infty$, then, whatever the initial distribution $\nu$,

$$\frac{1}{n} \sum_{i=1}^{n} h(X_i) \longrightarrow_{ps} \mathbb{E}_\Pi(h(X))$$
Theorem If the Markov chain $(X_k)_{k \in \mathbb{N}}$ is ergodic with stationary distribution $\Pi$ and if $h$ is a real function such that $\mathbb{E}_\Pi(|h(X)|) < \infty$, then, whatever the initial distribution $\nu$,

$$\frac{1}{n} \sum_{i=1}^{n} h(X_i) \rightarrow_{ps} \mathbb{E}_\Pi(h(X))$$

Convergence speed?
Definition The Markov chain \((X_k)_{k \in \mathbb{N}}\) with kernel \(P\) is said to be uniformly ergodic if there is \(M > 0\) and \(0 < r < 1\) such that

\[
\sup_{x \in \Omega} \sup_{A \in \mathcal{A}} |P^n(x, A) - \Pi(A)| \leq M r^n
\]
Convergence of Markov chains

**Definition** The Markov chain \((X_k)_{k \in \mathbb{N}}\) with kernel \(P\) is said to be uniformly ergodic if there is \(M > 0\) and \(0 < r < 1\) such that

\[
\sup_{x \in \Omega} \sup_{A \in \mathcal{A}} |P^n(x, A) - \Pi(A)| \leq Mr^n
\]

**Theorem** If the Markov chain \((X_k)_{k \in \mathbb{N}}\) is uniformly ergodic with stationary distribution \(\Pi\) and if \(h\) such that \(\mathbb{E}_\Pi(|h(X)|) < \infty\) then, whatever the initial distribution \(\nu\), there is \(\sigma(h) > 0\) such that

\[
\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} h(X_i) - \mathbb{E}_\Pi(h(X)) \right) \xrightarrow{d} \mathcal{N}(0, (\sigma(h))^2)
\]
The Metropolis-Hastings algorithm

Target distribution

\[ \Pi(dx) = \pi(x)\mu(dx) \]
The Metropolis-Hastings algorithm

Target distribution

\[ \Pi(dx) = \pi(x) \mu(dx) \]

Kernel \( Q \) for \( x \) such that \( \pi(x) > 0 \)

\[ Q(x, dy) = q(x, y) \mu(dy) \]
The Metropolis-Hastings algorithm

Choose $x^{(0)}$ such that $\pi(x^{(0)}) > 0$ and set $t = 1$

(∗) Generate $\tilde{x} \sim Q(x^{(t-1)}, \cdot)$

If $\pi(\tilde{x}) = 0$ then set $x^{(t)} = x^{(t-1)}$, $t = t + 1$ and return to (∗)

If $\pi(\tilde{x}) > 0$ calculate

$$\rho(x^{(t-1)}, \tilde{x}) = \frac{\pi(\tilde{x})/q(x^{(t-1)}, \tilde{x})}{\pi(x^{(t-1)})/q(\tilde{x}, x^{(t-1)})}$$

Generate $u \sim \mathcal{U}([0, 1])$

If $u \leq \rho(x^{(t-1)}, \tilde{x})$ then $x^{(t)} = \tilde{x}$ else $x^{(t)} = x^{(t-1)}$

set $t = t + 1$ and return to (∗)
The Metropolis-Hastings algorithm

Starting from $x$ ($\pi(x) > 0$), the acceptance probability of $y$ ($\pi(y) > 0$) is given by

$$\alpha(x, y) = \min \left[ 1, \frac{\pi(y)q(x, y)}{\pi(x)q(y, x)} \right]$$
The Metropolis-Hastings algorithm

Starting from $\pi(x) > 0$, the acceptance probability of $y$ ($\pi(y) > 0$) is given by

$$\alpha(x, y) = \min \left[ 1, \frac{\pi(y)/q(x, y)}{\pi(x)/q(y, x)} \right]$$

Whatever the value of $x$ such as $\pi(x) > 0$, the kernel associated with the Metropolis-Hastings algorithm is given by

$$K(x, dy) = q(x, y)\mu(dy)\alpha(x, y) + \left[ 1 - \int q(x, z)\alpha(x, z)\mu(dz) \right] \delta_x(dy)$$

where $\delta_x(\cdot)$ is the Dirac mass at point $x$
The Metropolis-Hastings algorithm

We can easily show that $K$ is $\Pi$-reversible

Indeed

$$\Pi(dx)K(x, dy) = \min [\pi(y)q(y, x), \pi(x)q(x, y)] \mu(dy)\mu(dx)$$

$$+ \left\{ \pi(x)\mu(dx) - \int \min [\pi(z)q(z, x), \pi(x)q(x, z)] \mu(dz) \right\} \delta_x(dy)$$

and

$$\Pi(dy)K(y, dx) = \min [\pi(x)q(x, y), \pi(y)q(y, x)] \mu(dx)\mu(dy)$$

$$+ \left\{ \pi(y)\mu(dy) - \int \min [\pi(x)q(x, z), \pi(z)q(z, x)] \mu(dz) \right\} \delta_y(dx)$$
The Metropolis-Hastings algorithm

**Theorem** If the kernel $Q$ is $\pi$-irreducible, the Markov chain generated with the Metropolis-Hastings algorithm is $\pi$-irreducible, $\pi$-invariant, Harris-recurrent and aperiodic.
The Metropolis-Hastings algorithm

**Theorem** If the kernel $Q$ is $\pi$-irreducible, the Markov chain generated with the Metropolis-Hastings algorithm is $\pi$-irreducible, $\pi$-invariant, Harris-recurrent and aperiodic

Two particular cases

- $Q$ is a random walk kernel: $q(x, y) = q_{RW}(x - y)$ and $q_{RW}(x) = q_{RW}(-x)$
- $Q$ is an independent kernel: $q(x, y) = q(y)$
The Gibbs sampler

Goal: generate simulations from multivariate distributions

Let \( X = (X_1, X_2, \ldots, X_d) \) with probability distribution \( \Pi \)

Note \( \Pi_i \) the conditional distribution of \( X_i \) given \( X_{-i} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_d) = x_{-i} \)

\( \Pi_i \) is called the full conditional distribution of \( X_i \)
The Gibbs sampler

Choose $x^{(0)}$ and set $t = 1$

$\ast$ Generate $x_1^{(t)} \sim \Pi_1(\cdot | x_2^{(t-1)}, \ldots, x_d^{(t-1)})$

Generate $x_2^{(t)} \sim \Pi_2(\cdot | x_1^{(t)}, x_3^{(t-1)}, \ldots, x_d^{(t-1)})$

Generate $x_3^{(t)} \sim \Pi_3(\cdot | x_1^{(t)}, x_2^{(t)}, x_4^{(t-1)} \ldots, x_d^{(t-1)})$

$\ldots$

Generate $x_d^{(t)} \sim \Pi_d(\cdot | x_1^{(t)}, \ldots, x_{d-1}^{(t)})$

Set $t = t + 1$ and return to $\ast$
The Gibbs sampler

Choose \( x^{(0)} \) and set \( t = 1 \)

\((*)\) Generate \( x_1^{(t)} \sim \Pi_1( \cdot | x_2^{(t-1)}, \ldots, x_d^{(t-1)} ) \)

Generate \( x_2^{(t)} \sim \Pi_2( \cdot | x_1^{(t)}, x_3^{(t-1)}, \ldots, x_d^{(t-1)} ) \)

Generate \( x_3^{(t)} \sim \Pi_3( \cdot | x_1^{(t)}, x_2^{(t)}, x_4^{(t-1)} \ldots, x_d^{(t-1)} ) \)

\[ \ldots \]

Generate \( x_d^{(t)} \sim \Pi_d( \cdot | x_1^{(t)}, \ldots, x_{d-1}^{(t)} ) \)

Set \( t = t + 1 \) and return to \((*)\)

**Theorem** The Markov chain generated using the Gibbs sampler is \( \Pi \)-irreducible, \( \Pi \)-invariant, Harris-recurrent and aperiodic