The Cut-off phenomenon for Monte Carlo Markov Chains

IRTG Stochastic Models of Complex Processes

Michele Salvi

Roma3 University

February 16, 2010

Reversible Markov Chains

Consider a discrete-time Markov Chain on a finite state space Ω with transition matrix P that is

Reversible Markov Chains

Consider a discrete-time Markov Chain on a finite state space Ω with transition matrix P that is

- irreducible;

Reversible Markov Chains

Consider a discrete-time Markov Chain on a finite state space Ω with transition matrix P that is

- irreducible;
- aperiodic;

Reversible Markov Chains

Consider a discrete-time Markov Chain on a finite state space Ω with transition matrix P that is

- irreducible;
- aperiodic;
- reversible with respect to a probability measure π, that is

$$\pi(x)P(x,y) = \pi(y)P(y,x), \qquad \forall x, y \in \Omega. \tag{1}$$

Reversible Markov Chains

Consider a discrete-time Markov Chain on a finite state space Ω with transition matrix P that is

- irreducible;
- aperiodic;
- reversible with respect to a probability measure π, that is

$$\pi(x)P(x,y) = \pi(y)P(y,x), \qquad \forall x, y \in \Omega. \tag{1}$$

If π satisfies (1), known as *detailed balance equations*, then it is a stationary distribution for our chain P:

$$\pi P = \pi.$$

Theorem (Ergodic theorem)

Let (X_t) be an irreducible and aperiodic Markov chain with transition matrix P. Then there exists a unique stationary distribution π for the chain. Moreover,

$$P^t(x,y) \xrightarrow{t\to\infty} \pi(y) \qquad \forall x,y \in \Omega.$$

Theorem (Ergodic theorem)

Let (X_t) be an irreducible and aperiodic Markov chain with transition matrix P. Then there exists a unique stationary distribution π for the chain. Moreover,

$$P^t(x,y) \xrightarrow{t\to\infty} \pi(y) \qquad \forall x,y \in \Omega.$$

?

Problem of the approach to the equilibrium:
how *fast* is this convergence?

We have first to define a notion of *distance*.
The total variation distance between two probability measures μ and ν is

$$\|\mu - \nu\|_{TV} := \max_{A \subset \Omega} |\mu(A) - \nu(A)|$$

We have first to define a notion of *distance*.

The total variation distance between two probability measures μ and ν is

$$\|\mu - \nu\|_{TV} := \max_{A \subset \Omega} |\mu(A) - \nu(A)| = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|.$$

We have first to define a notion of *distance*.
The total variation distance between two probability measures μ and ν is

$$\|\mu - \nu\|_{TV} := \max_{A \subset \Omega} |\mu(A) - \nu(A)| = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|.$$

Then we can define the distance of a Markov Chain P from its stationary distribution π at step t as

$$d(t) := \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{TV}. \tag{2}$$

It is easy to show that this is a decreasing function of t.

Mixing time

The ε-mixing time of a Markov Chain is

$$t_{mix}(\varepsilon) := \min\{t : d(t) \leq \varepsilon\}. \tag{3}$$

We also set for simplicity $t_{mix} := t_{mix}(1/4)$.

One possible approach to the study of the mixing time is the study of the eigenvalues of P. Let's label them in decreasing order:

$$1 = \lambda_1 \geq \lambda_2 \geq ... \geq \lambda_{|\Omega|} \geq -1.$$

One possible approach to the study of the mixing time is the study of the eigenvalues of P. Let's label them in decreasing order:

$$1 = \lambda_1 \geq \lambda_2 \geq ... \geq \lambda_{|\Omega|} \geq -1.$$

- P is irreducible $\implies$ $\lambda_2 < 1$;

One possible approach to the study of the mixing time is the study of the eigenvalues of P. Let's label them in decreasing order:

$$1 = \lambda_1 \geq \lambda_2 \geq ... \geq \lambda_{|\Omega|} \geq -1.$$

- P is irreducible $\Longrightarrow \lambda_2 < 1$;
- P is lazy $\Longrightarrow \lambda_i \geq 0, \forall i$.

One possible approach to the study of the mixing time is the study of the eigenvalues of P. Let's label them in decreasing order:

$$1 = \lambda_1 \geq \lambda_2 \geq ... \geq \lambda_{|\Omega|} \geq -1.$$

- P is irreducible $\implies \lambda_2 < 1$;
- P is lazy $\implies \lambda_i \geq 0, \forall i$.

Let $\mathrm{gap} := 1 - \lambda_2$ be the spectral gap of the chain and call its inverse $t_{rel} := \frac{1}{\mathrm{gap}}$ the relaxation time of the chain. Then

One possible approach to the study of the mixing time is the study of the eigenvalues of P. Let's label them in decreasing order:

$$1 = \lambda_1 \geq \lambda_2 \geq ... \geq \lambda_{|\Omega|} \geq -1.$$

- P is irreducible $\implies \lambda_2 < 1$;
- P is lazy $\implies \lambda_i \geq 0, \ \forall i$.

Let $\mathrm{gap} := 1 - \lambda_2$ be the spectral gap of the chain and call its inverse $t_{rel} := \frac{1}{\mathrm{gap}}$ the relaxation time of the chain. Then

Proposition

$$(t_{rel} - 1)\log\left(\frac{1}{2\varepsilon}\right) \leq t_{mix}(\varepsilon) \leq \log\left(\frac{1}{\varepsilon\pi_{min}}\right) t_{rel}. \tag{4}$$

Cut-off

Definition

Let $(X_t^{(n)})$ be a sequence of Markov Chains on state spaces $\Omega^{(n)}$, with transition matrices $P^{(n)}$, stationary distributions $\pi^{(n)}$ and ε-mixing times $t_{mix}^{(n)}(\varepsilon)$. We say that this sequence exhibits Cut-off if

Cut-off

Definition

Let $(X_t^{(n)})$ be a sequence of Markov Chains on state spaces $\Omega^{(n)}$, with transition matrices $P^{(n)}$, stationary distributions $\pi^{(n)}$ and ε-mixing times $t_{mix}^{(n)}(\varepsilon)$. We say that this sequence exhibits Cut-off if

$$\lim_{n\to\infty} \frac{t_{mix}^{(n)}(\varepsilon)}{t_{mix}^{(n)}(1-\varepsilon)} = 1, \qquad \forall 0 \le \varepsilon \le \frac{1}{2}. \tag{5}$$

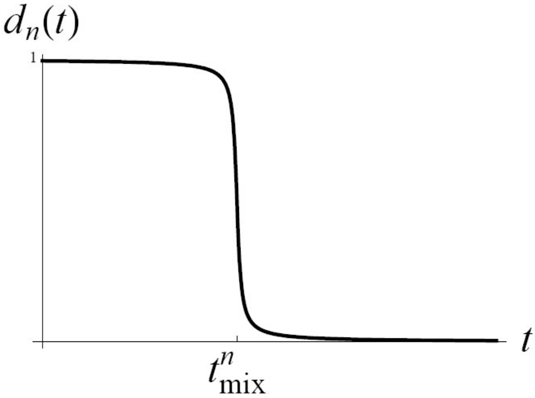

Figure: *For a sequence of chains with Cut-off, the graph of $d_n(t)$, zoomed on a time-scale of $t^{(n)}_{mix}$, approaches a step function as $n \to \infty$.*

Cut-off window

Definition

We say that a sequence of Markov Chains has Cut-off with a window of size ω_n *if*

Cut-off window

Definition

We say that a sequence of Markov Chains has Cut-off with a window of size ω_n *if*

- $\omega_n = o(t_{mix}^{(n)})$

Cut-off window

Definition

We say that a sequence of Markov Chains has Cut-off with a window of size ω_n if

- $\omega_n = o(t_{mix}^{(n)})$
- $\lim_{\alpha\to\infty} \liminf_{n\to\infty} d_n(t_{mix}^{(n)} - \alpha\omega_n) = 1,$

Cut-off window

Definition

We say that a sequence of Markov Chains has Cut-off with a window of size ω_n if

- $\omega_n = o(t_{mix}^{(n)})$
- $\lim_{\alpha\to\infty} \liminf_{n\to\infty} d_n(t_{mix}^{(n)} - \alpha\omega_n) = 1,$
- $\lim_{\alpha\to\infty} \limsup_{n\to\infty} d_n(t_{mix}^{(n)} + \alpha\omega_n) = 0.$

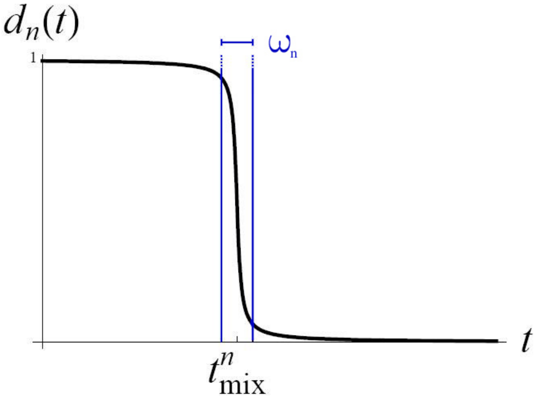

Figure: *In a small time-interval around* $t_{mix}^{(n)}$, *the distance falls from near* 1 *to almost* 0.

Examples

Consider a graph $G = (V, E)$; in each vertex we can place either a positive or a negative spin.

Examples

Consider a graph $G = (V, E)$; in each vertex we can place either a positive or a negative spin. The state space of our process is $\Omega = \{-1, +1\}^V$.

Examples

Consider a graph $G = (V, E)$; in each vertex we can place either a positive or a negative spin. The state space of our process is $\Omega = \{-1, +1\}^V$.
We assign to every configuration $\sigma \in \Omega$ an energy given by

$$H(\sigma) := -\frac{1}{n} \sum_{\substack{v,w \in V \\ v \sim w}} \sigma(v)\sigma(w).$$

Examples

Consider a graph $G = (V, E)$; in each vertex we can place either a positive or a negative spin. The state space of our process is $\Omega = \{-1, +1\}^V$.
We assign to every configuration $\sigma \in \Omega$ an energy given by

$$H(\sigma) := -\frac{1}{n} \sum_{\substack{v,w \in V \\ v \sim w}} \sigma(v)\sigma(w).$$

Then we can define the Gibbs measure on Ω given by

$$\mu(\sigma) := \frac{e^{-\beta H(\sigma)}}{Z(\beta)},$$

where $Z(\beta)$ is the partition function and the parameter β can be interpreted as the inverse of the temperature.

The heat-bath Glauber dynamics on this model is the following:

The heat-bath Glauber dynamics on this model is the following:

- select uniformly at random a vertex in V;

The heat-bath Glauber dynamics on this model is the following:

- select uniformly at random a vertex in V;
- 'erase' its spin;

The heat-bath Glauber dynamics on this model is the following:

- select uniformly at random a vertex in V;
- 'erase' its spin;
- replace it with a brand new spin generated according to μ.

The heat-bath Glauber dynamics on this model is the following:

- select uniformly at random a vertex in V;
- 'erase' its spin;
- replace it with a brand new spin generated according to μ.

That is, if we choose vertex $w \in V$ to be updated in a configuration $\sigma \in \Omega$, the probability of putting a spin $x \in \pm 1$ in w is

$$p(\sigma, x) = \frac{e^{\beta S(\sigma,x)}}{e^{\beta S(\sigma,x)} + e^{-\beta S(\sigma,x)}}, \tag{6}$$

where $S(\sigma, x) := \sum_{v \sim w} \sigma(v)$.

The heat-bath Glauber dynamics on this model is the following:

- select uniformly at random a vertex in V;
- 'erase' its spin;
- replace it with a brand new spin generated according to μ.

That is, if we choose vertex $w \in V$ to be updated in a configuration $\sigma \in \Omega$, the probability of putting a spin $x \in \pm 1$ in w is

$$p(\sigma, x) = \frac{e^{\beta S(\sigma,x)}}{e^{\beta S(\sigma,x)} + e^{-\beta S(\sigma,x)}}, \tag{6}$$

where $S(\sigma, x) := \sum_{v \sim w} \sigma(v)$.

This dynamics is reversible with respect to measure μ.

Theorem (Levin, Luczak, Peres, 2007)

Let $(X_t^{(n)})$ be the Glauber dynamics for the Ising model on the n-complete graph. If $\beta < 1$, there is a Cut-off at time $t_n := \frac{n \log n}{2(1-\beta)}$ with window $\omega_n = O(n)$.

Theorem (Levin, Luczak, Peres, 2007)

Let $(X_t^{(n)})$ be the Glauber dynamics for the Ising model on the n-complete graph. If $\beta < 1$, there is a Cut-off at time $t_n := \frac{n \log n}{2(1-\beta)}$ with window $\omega_n = O(n)$.
Furthermore, if $\beta = 1$, then $t_{mix} = O(n^{\frac{3}{2}})$, if $\beta > 1$, t_{mix} has an exponential behaviour, and in both case there is no Cut-off.

Main idea:

- The big symmetry of the model allows us to deal just with a simpler process, namely the magnetization chain:

$$M_t^{(n)} := \frac{1}{n}\sum_{v=1}^{n} X_t^{(n)}(v),$$

with state space $\Omega_M^{(n)} := \{-1, -1+\frac{2}{n}, ..., 1\}$ (B&D chain).

Main idea:

- The big symmetry of the model allows us to deal just with a simpler process, namely the magnetization chain:

$$M_t^{(n)} := \frac{1}{n} \sum_{v=1}^{n} X_t^{(n)}(v),$$

with state space $\Omega_M^{(n)} := \{-1, -1 + \frac{2}{n}, ..., 1\}$ (B&D chain).

- We use a coupling of the original dynamics to make the magnetization of the two copies of the chain merge in t_n steps (in particular the most of the work is to make them differ for less than $\frac{cost.}{\sqrt{n}}$, which is achieved via the monotone coupling strategy).

Main idea:

- The big symmetry of the model allows us to deal just with a simpler process, namely the magnetization chain:

$$M_t^{(n)} := \frac{1}{n}\sum_{v=1}^{n} X_t^{(n)}(v),$$

with state space $\Omega_M^{(n)} := \{-1, -1+\frac{2}{n}, ..., 1\}$ (B&D chain).

- We use a coupling of the original dynamics to make the magnetization of the two copies of the chain merge in t_n steps (in particular the most of the work is to make them differ for less than $\frac{cost.}{\sqrt{n}}$, which is achieved via the monotone coupling strategy).
- Once the magnetizations have met, we need only other $O(n)$ steps to make the two copies coincide.

Theorem (Lubetzky, Sly, 2009)

Let $(X_t^{(n)})$ be the continuous-time Glauber dynamics for the Ising model on the lattice $\left(\frac{\mathbb{Z}}{n\mathbb{Z}}\right)^2$ with periodic boundary conditions. Let $\beta_c = \frac{1}{2}\log(1+\sqrt{2})$ be the critical inverse-temperature.

Theorem (Lubetzky, Sly, 2009)

Let $(X_t^{(n)})$ be the continuous-time Glauber dynamics for the Ising model on the lattice $\left(\frac{\mathbb{Z}}{n\mathbb{Z}}\right)^2$ with periodic boundary conditions. Let $\beta_c = \frac{1}{2}\log(1+\sqrt{2})$ be the critical inverse-temperature.
Then, if $0 \leq \beta < \beta_c$, there is Cut-off at time $t_n := \lambda_\infty^{-1} \log n$ with a window of size $O(\log\log n)$, where λ_∞ is the spectral gap of the dynamics on the infinite-volume lattice.

Theorem (Lubetzky, Sly, 2009)

Let $(X_t^{(n)})$ be the continuous-time Glauber dynamics for the Ising model on the lattice $\left(\frac{\mathbb{Z}}{n\mathbb{Z}}\right)^2$ with periodic boundary conditions. Let $\beta_c = \frac{1}{2}\log(1+\sqrt{2})$ be the critical inverse-temperature. Then, if $0 \le \beta < \beta_c$, there is Cut-off at time $t_n := \lambda_\infty^{-1}\log n$ with a window of size $O(\log\log n)$, where λ_∞ is the spectral gap of the dynamics on the infinite-volume lattice.

The Theorem can be extended to the case of the $\left(\frac{\mathbb{Z}}{n\mathbb{Z}}\right)^d$, with $d \in \mathbb{N}$, whenever β and h (the external field) are such that the strong spatial mixing property holds.

Main idea:

- Fundamental concept of update support: the set of spins which, given an update sequence, actually influences the final configuration.

Main idea:

- Fundamental concept of update support: the set of spins which, given an update sequence, actually influences the final configuration.
- This set is shown to be sparse with high probability, that is, concentrated in small islands far enough from each other to be considered almost independent.

Main idea:

- Fundamental concept of update support: the set of spins which, given an update sequence, actually influences the final configuration.
- This set is shown to be sparse with high probability, that is, concentrated in small islands far enough from each other to be considered almost independent.
- Studying the L^2-mixing of these smaller lattices is enough to control the mixing of the whole dynamics.

- 1981: First proof of the existence of a Cut-off for the random transposition Markov chain on the symmetric group by Diaconis and Shahshahani (never used the term 'Cut-off');

- 1981: First proof of the existence of a Cut-off for the random transposition Markov chain on the symmetric group by Diaconis and Shahshahani (never used the term 'Cut-off');
- 1986: Introduction of the notion of Cut-off, proof of its existence for the top-in-at-random card shuffling chain by Diaconis and Aldous;

- 1981: First proof of the existence of a Cut-off for the random transposition Markov chain on the symmetric group by Diaconis and Shahshahani (never used the term 'Cut-off');
- 1986: Introduction of the notion of Cut-off, proof of its existence for the top-in-at-random card shuffling chain by Diaconis and Aldous;
- 1995: in his survey of the Cut-off phenomenon Diaconis says:
  "*At present writing, proof of a cutoff is a diffcult, delicate affair, requiring detailed knowledge of the chain, such as all eigenvalues and eigenvectors. Most of the examples where this can be pushed through arise from random walk on groups, with the walk having a fair amount of symmetry.*";

- 1981: First proof of the existence of a Cut-off for the random transposition Markov chain on the symmetric group by Diaconis and Shahshahani (never used the term 'Cut-off');
- 1986: Introduction of the notion of Cut-off, proof of its existence for the top-in-at-random card shuffling chain by Diaconis and Aldous;
- 1995: in his survey of the Cut-off phenomenon Diaconis says:
  "*At present writing, proof of a cutoff is a diffcult, delicate affair, requiring detailed knowledge of the chain, such as all eigenvalues and eigenvectors. Most of the examples where this can be pushed through arise from random walk on groups, with the walk having a fair amount of symmetry.*";
- 2004: Cut-off proved for Random Walks on finite groups (Saloff-Coste);

- 1981: First proof of the existence of a Cut-off for the random transposition Markov chain on the symmetric group by Diaconis and Shahshahani (never used the term 'Cut-off');
- 1986: Introduction of the notion of Cut-off, proof of its existence for the top-in-at-random card shuffling chain by Diaconis and Aldous;
- 1995: in his survey of the Cut-off phenomenon Diaconis says:
  "*At present writing, proof of a cutoff is a diffcult, delicate affair, requiring detailed knowledge of the chain, such as all eigenvalues and eigenvectors. Most of the examples where this can be pushed through arise from random walk on groups, with the walk having a fair amount of symmetry.*";
- 2004: Cut-off proved for Random Walks on finite groups (Saloff-Coste);
- 2004: Peres conjectures that a necessary and sufficient condition for Cut-off is that

$$\mathrm{gap}^{(n)} \cdot t_{mix}^{(n)} \xrightarrow{n\to\infty} \infty. \tag{7}$$

  This was shown not to be always true, but is still believed to hold for many 'natural' classes of Markov Chains (e.g. proved for B&D Markov Chains in 2008);

- 1981: First proof of the existence of a Cut-off for the random transposition Markov chain on the symmetric group by Diaconis and Shahshahani (never used the term 'Cut-off');
- 1986: Introduction of the notion of Cut-off, proof of its existence for the top-in-at-random card shuffling chain by Diaconis and Aldous;
- 1995: in his survey of the Cut-off phenomenon Diaconis says: "*At present writing, proof of a cutoff is a diffcult, delicate affair, requiring detailed knowledge of the chain, such as all eigenvalues and eigenvectors. Most of the examples where this can be pushed through arise from random walk on groups, with the walk having a fair amount of symmetry.*";
- 2004: Cut-off proved for Random Walks on finite groups (Saloff-Coste);
- 2004: Peres conjectures that a necessary and sufficient condition for Cut-off is that

$$\mathrm{gap}^{(n)} \cdot t_{mix}^{(n)} \xrightarrow{n\to\infty} \infty. \tag{7}$$

  This was shown not to be always true, but is still believed to hold for many 'natural' classes of Markov Chains (e.g. proved for B&D Markov Chains in 2008);
- 2009: Cut-off proved for the Glauber dynamics for the Ising model on the lattice in any dimension, whenever strong spatial mixing holds, by Sly and Lubetzky.

Bibliography

Aldous D. and Diaconis P. (1986) *Shuffling cards and stopping times*, Amer. Math. Monthly 93, 333-348.

Aldous D. and Fill J. (1999) *Reversible Markov chains and random walks on graphs*, in progress. Manuscript available at http://www.stat.berkeley.edu/∼aldous/RWG/book.html.

Chen G.-Y. and Saloff-Coste L. (2008) *The cutoff phenomenon for ergodic Markov processes*, Electronic Journal of Probability, Vol. 13, paper no. 3, 26-78, http://www.math.washington.edu/∼ejpecp.

Ding J., Lubetzky E. and Peres Y. (2008) *The mixing time evolution of Glauber dynamics for the mean-field Ising model*, arXiv:0806.1906v2.

Ding J., Lubetzky E. and Peres Y. (2008) *Total variation cutoff in birth and death chains*, arXiv:0801.2625v4.

Levin D. A., Malwina L. J. and Peres Y. (2007) *Glauber dynamics for the mean-field Ising model: cut-off, critical power law, and metastability*, arXiv:0712.0790v2.

Levin D. A., Peres Y. and Wilmer E. L. (2009) *Markov Chains and Mixing Times*, American Mathematical Society.

Lubetzky E. and Sly A. (2009) *Cutoff for the Ising model on the lattice*, arXiv:0909.4320v1.

Saloff-Coste L. (1997) *Lectures on finite Markov chains*, Lectures On Probability Theory and Statistics, École d'Ete de Probabilites de Saint-Flour XXVI, 301-413.

Thank you!