



UNIVERSITÀ DEGLI STUDI ROMA TRE
FACOLTÀ DI SCIENZE M.F.N.
CORSO DI LAUREA IN MATEMATICA

Tesi di Laurea Magistrale in Matematica

**The Cut-off Phenomenon
for Monte Carlo Markov Chains**
(una sintesi)

Candidato
Michele Salvi

Relatore
Prof. Fabio Martinelli

ANNO ACCADEMICO 2008-2009
Ottobre 2009

Classificazione AMS: 60J10, 82B20, 82B26, 82C20, 60K35, 82C22, 82C80.
Parole chiave: Markov chains, cut-off, mixing time, Ising model, Glauber dynamics.

Introduzione

L'argomento principale della tesi "The Cut-off phenomenon for Monte Carlo Markov chains" riguarda le catene di Markov reversibili rispetto ad una misura di probabilità π su uno spazio degli stati finito Ω . La varietà di campi scientifici in cui esse trovano applicazione, dalla genetica all'informatica teorica, dalla fisica statistica all'economia, dalla crittografia fino alle scienze sociali, giustifica il fervore della ricerca intorno a questo argomento e la sua fertilità. Di particolare interesse è il loro impiego nell'ambito della simulazione: con il metodo Monte Carlo è possibile campionare con buona approssimazione da misure di probabilità su spazi con alta cardinalità, in situazioni che sarebbero altrimenti pressochè intrattabili (anche al calcolatore) con metodi deterministici.

La teoria classica per le catene di Markov studiava la convergenza di una catena *fissata* a π , unica misura reversibile e stazionaria, ed il suo tempo di mescolamento t_{mix} . Il passo successivo è quello di prendere in considerazione *famiglie* di catene $(X^{(n)}, \pi^{(n)})$, parametrizzate ad esempio dalla taglia n dello spazio degli stati, e studiare l'andamento asintotico del loro tempo di mescolamento $t_{mix}^{(n)}$ in funzione di n .

In particolare alcune tra le famiglie di catene più studiate presentano un interessante comportamento comune: per un tempo dell'ordine di $t_{mix}^{(n)}$ esse si mantengono quasi alla distanza (per esempio in variazione totale) massima da $\pi^{(n)}$; poi, in un intervallo temporale $\omega^{(n)}$ relativamente trascurabile ($\omega^{(n)} = o(t_{mix}^{(n)})$), questa distanza si riduce drasticamente fino quasi a 0. Tuttavia la dimostrazione rigorosa dell'esistenza di questo fenomeno, noto come *Cut-off phenomenon*, risulta essere estremamente difficile nella maggior parte dei casi.

La tesi si ripropone, dopo una prima parte introduttiva, di esporre alcuni tra i risultati recenti più rilevanti sull'argomento. Nell'ultima parte del lavoro vengono fornite stime analitiche di alcune fondamentali quantità nello studio del rilassamento all'equilibrio (tra cui la costante logaritmica di Sobolev) per uno dei modelli per i quali è stata provata l'esistenza del Cut-off, la dinamica di Glauber per il modello di Ising su grafo completo ad alta temperatura.

1 Catene di Markov su spazi finiti

Una **catena di Markov** su uno spazio finito Ω è un processo stocastico a tempo discreto (X_t) descritto da una matrice stocastica $|\Omega| \times |\Omega|$, sia P , che rispetta la cosiddetta **proprietà di Markov**: $\forall t > 0$ and $\forall x, y \in \Omega$

$$\begin{aligned} \mathbb{P}(X_{t+1} = y | X_0 = x_0, \dots, X_{t-1} = x_{t-1}, X_t = x) &= \mathbb{P}(X_{t+1} = y | X_t = x) \\ &= P_{xy}, \end{aligned} \quad (1)$$

dove x_0, \dots, x_{t-1} è una qualsiasi sequenza di punti di Ω . Data una distribuzione iniziale μ_0 su Ω , la probabilità di trovarsi nello stato y al tempo t è data da

$$(\mu_0 P^t)(y) = \sum_{x_0 \in \Omega} \dots \sum_{x_{t-1} \in \Omega} \mu_0(x_0) P(x_0, x_1) \dots P(x_{t-1}, y).$$

Se $\mu_0 = \delta_x$, ovvero nel caso in cui si parta certamente da uno stato x , allora la probabilità di trovarsi in y dopo t passi è semplicemente $P^t(x, y)$.

Una catena di Markov viene detta **irriducibile** se, $\forall x, y \in \Omega$, esiste $t \geq 0$ (che può dipendere da x e y stessi) tale che $P^t(x, y) > 0$. Sia ora $T(x) := \{t \geq 1 : P^t(x, x) > 0\}$ l'insieme dei tempi in cui, partendo da $x \in \Omega$, è possibile tornare in x stesso. Se il gran comun divisore di $T(\cdot)$ è 1 per tutti gli stati in Ω , allora la catena è detta **aperiodica**.

Una catena Q non aperiodica può comunque essere facilmente modificata affinché lo diventi: basta definire il nuovo processo con matrice di transizione $Q' := \frac{1}{2}Q + \frac{1}{2}Id$, dove Id è la matrice identità. E' facile verificare che Q' , detta la versione **pigra** di Q , non solo è aperiodica, ma conserva anche molte delle caratteristiche della catena originaria.

Una misura di probabilità π su Ω è detta **stazionaria** per una catena di Markov P se verifica:

$$\pi P = \pi. \quad (2)$$

Il seguente teorema, noto come **Teorema Ergodico**, dimostra l'importanza di queste distribuzioni:

Teorema 1.1. *Sia $\{X_t\}$ una catena di Markov irriducibile ed aperiodica con matrice di transizione P . Allora esiste un'unica misura stazionaria π per la catena. Inoltre*

$$P^t(x, y) \xrightarrow{t \rightarrow \infty} \pi(y) \quad \forall x, y \in \Omega. \quad (3)$$

Altra importante nozione è quella di **reversibilità**: una catena di Markov P è detta reversibile rispetto ad una misura di probabilità π se, $\forall x, y \in \Omega$,

$$\pi(x)P(x, y) = \pi(y)P(y, x). \quad (4)$$

L'equazione (4) è spesso chiamata **bilancio dettagliato**. Se una distribuzione verifica il bilancio dettagliato, allora è stazionaria per la catena data. Verificare (4) è spesso la maniera più semplice per capire se una misura è stazionaria per una certa catena.

Sono definibili diverse metriche sullo spazio delle misure di probabilità su Ω . In questo lavoro tratteremo quasi unicamente la distanza in **variazione totale**: se μ e ν sono due misure di probabilità su Ω , la loro distanza è

$$\|\mu - \nu\|_{TV} := \max_{A \in \Omega} |\mu(A) - \nu(A)| = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|. \quad (5)$$

A questo punto è possibile definire la distanza della nostra catena P dalla sua misura stazionaria π al tempo t :

$$d(t) := \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{TV}. \quad (6)$$

$d(\cdot)$ è una funzione decrescente ed inoltre si ha, come conseguenza del Teorema Ergodico, che se P è irriducibile ed aperiodica, allora esistono costanti $\alpha \in (0, 1)$ e $c > 0$ tali che

$$\max_{x \in \Omega} \|P^t(x, \cdot) - \pi(\cdot)\|_{TV} \leq C\alpha^t. \quad (7)$$

La convergenza alla misura stazionaria misurata in variazione totale è cioè di tipo esponenziale.

Il primo tempo in cui la distanza dalla misura stazionaria è minore di un certo $\varepsilon \in (0, 1)$ è detto ε -tempo di mixing della catena:

$$t_{mix}(\varepsilon) := \min \{t : d(t) \leq \varepsilon\}. \quad (8)$$

Per convenzione $t_{mix} := t_{mix}(1/4)$ viene chiamato semplicemente **tempo di mixing**.

La seconda parte del primo capitolo descrive alcune tecniche per la stima del tempo di mixing di una catena. Il primo è lo studio spettrale della matrice di transizione P . Sia $|\Omega| = n$. Etichettiamo gli n autovalori di P in ordine decrescente (tutti minori di 1 in modulo poichè P è stocastica):

$$1 = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{|\Omega|} \geq -1. \quad (9)$$

Se P è irriducibile, allora $\lambda_2 < 1$, mentre se la catena è pigra, allora non ci sono autovalori negativi. In questo caso chiamiamo $\gamma := 1 - \lambda_2$ lo **spectral gap** della catena e

$$t_{rel} := \frac{1}{\gamma} \quad (10)$$

il suo **tempo di rilassamento**. Questa quantità è strettamente legata al tempo di mixing, infatti, se la catena è reversibile, irriducibile ed aperiodica,

$$(t_{rel} - 1) \log \left(\frac{1}{2\varepsilon} \right) \leq t_{mix}(\varepsilon) \leq \log \left(\frac{1}{\varepsilon \pi_{min}} \right) t_{rel}, \quad (11)$$

dove $\pi_{min} := \min_{x \in \Omega} \pi(x)$.

La seconda tecnica è quella del **coupling**. Un coupling di due misure di probabilità μ e ν su uno spazio Ω è una coppia di variabili aleatorie (X, Y) le cui distribuzioni marginali sono rispettivamente μ e ν . Forzando il più possibile X ed Y ad assumere lo stesso valore, si può dimostrare il seguente utile risultato:

Proposizione 1.2. *Date due misure di probabilità μ e ν su Ω , si ha che*

$$\|\mu - \nu\|_{TV} = \inf \{ \mathbb{P}(X \neq Y) : (X, Y) \text{ è coupling di } \mu \text{ e } \nu \}. \quad (12)$$

Analogamente possiamo definire il coupling di due catene di Markov su Ω che partono da diverse distribuzioni μ_0 e ν_0 , ma che evolvono secondo la stessa matrice di transizione P ; il loro coupling sarà un processo $(X_t, Y_t)_{t \geq 0}$ tale che $X_t \sim P_{\mu_0}^t$ e $Y_t \sim P_{\nu_0}^t$. Il seguente utilissimo risultato è conseguenza della Proposizione 1.2:

Corollario 1.3. *Supponiamo che per ogni coppia di stati $x, y \in \Omega$ ci sia un coupling $(X_t, Y_t)_{t \geq 0}$ tale che $X_0 = x, Y_0 = y$. Allora*

$$d(t) \leq \max_{x, y \in \Omega} \mathbb{P}_{x, y}(\tau_c > t), \quad (13)$$

dove $\mathbb{P}_{x, y}(\cdot)$ è la probabilità sullo spazio in cui sono definite entrambe le catene e $\tau_c := \min\{t : X_t = Y_t\}$.

2 Il fenomeno del Cut-off

Consideriamo una sequenza di catene di Markov indicizzate da $n = 1, 2, \dots$ con i loro spazi degli stati $\Omega^{(n)}$; si può pensare che, al crescere di n , cresca la taglia degli Ω . Supponiamo che ognuna di queste catene abbia la sua distribuzione stazionaria $\pi^{(n)}$ e il suo tempo di mixing $t_{mix}^{(n)}(\varepsilon)$. Diciamo che la sequenza presenta un **cut-off** se, per ogni $0 < \varepsilon < \frac{1}{2}$,

$$\lim_{n \rightarrow \infty} \frac{t_{mix}^{(n)}(\varepsilon)}{t_{mix}^{(n)}(1 - \varepsilon)} = 1. \quad (14)$$

Graficamente, se disegniamo il grafico di $d_n(t)$ (ovvero la $d(t)$ corrispondente all' n -esima catena) e riscaldiamo la figura nell'ordine di $t_{mix}^{(n)}$, vediamo che al crescere di n la funzione approssima sempre più una funzione a gradino con salto in $t_{mix}^{(n)}$ (Figura nella prossima pagina).

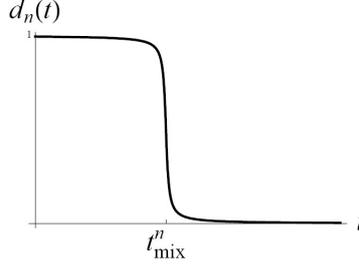


Figure 1: Per una sequenza di catene con cut-off, il grafico di $d_n(t)$ in funzione di t , visto su una scala di $t_{mix}^{(n)}$, approssima una funzione a gradino quando $n \rightarrow \infty$.

Una definizione equivalente è: una sequenza di catene di Markov presenta un cut-off se e solo se

$$d_n(c t_{mix}^{(n)}) \xrightarrow{n \rightarrow \infty} \begin{cases} 1 & \text{if } c < 1 \\ 0 & \text{if } c > 1. \end{cases} \quad (15)$$

Conoscere l'esistenza del cut-off per una famiglia di catene di Markov può essere estremamente utile: se vogliamo approssimare la distribuzione $\pi^{(n)}$ (per esempio in una simulazione), sappiamo di dover aspettare almeno $t_{mix}^{(n)}$ passi della catena, ma sappiamo anche che attendere molto di più risulta abbastanza inutile. Questa interessante proprietà sembra presentarsi in molti dei processi più studiati; essendo però estremamente difficile verificare che una sequenza soddisfi la definizione, al giorno d'oggi il numero di catene per cui si è dimostrata rigorosamente l'esistenza di cut-off è estremamente ristretta.

Per specificare ulteriormente il carattere di questa convergenza introduciamo la nozione di **finestra del cut-off**. Diciamo che una sequenza di catene di Markov ha un cut-off con finestra di ordine ω_n se $\omega_n = o(t_{mix}^{(n)})$ e

$$\lim_{\alpha \rightarrow \infty} \liminf_{n \rightarrow \infty} d_n(t_{mix}^{(n)} - \alpha \omega_n) = 1, \quad (16)$$

$$\lim_{\alpha \rightarrow \infty} \limsup_{n \rightarrow \infty} d_n(t_{mix}^{(n)} + \alpha \omega_n) = 0. \quad (17)$$

ω_n è sostanzialmente l'ordine di grandezza dell'intervallo intorno a $t_{mix}^{(n)}$ in cui la distanza della catena dalla sua misura stazionaria crolla da quasi 1 (il massimo) a quasi 0.

Una condizione necessaria affinché si verifichi il cut-off in una sequenza di catene di Markov irriducibili, aperiodiche e reversibili, con gap spettrali $\text{gap}^{(n)}$, è che

$$\lim_{n \rightarrow \infty} t_{mix}^{(n)} \cdot \text{gap}^{(n)} = \infty. \quad (18)$$

Nella seconda parte di questo capitolo viene esibito il primo esempio concreto. L'ipercono n -dimensionale è un grafo i cui vertici sono le stringhe

binarie di lunghezza n ($\Omega = \{0, 1\}^n$); due punti di Ω sono connessi da un arco se e solo se differiscono per una sola coordinata. La nostra catena di Markov (X_t) sarà la **camminata aleatoria semplice sull'ipercubo**: stando su di un vertice \bar{x} , si sceglierà la prossima posizione uniformemente tra tutti gli n vertici connessi al proprio tramite un arco. Ovvero, ad ogni passo della catena si sceglierà con probabilità $\frac{1}{n}$ una delle n possibili coordinate e le si cambierà di valore. La versione pigra di questo processo prevede la possibilità di rimanere ad ogni passo nella posizione corrente con probabilità $\frac{1}{2}$. In entrambi i casi la misura stazionaria π è quella uniforme su tutti i 2^n vertici.

Teorema 2.1. *Per la camminata aleatoria (pigra) semplice sull'ipercubo n -dimensionale si ha un cut-off al tempo $t_n := \frac{1}{2}n \log n$ con finestra di ordine n .*

Per dimostrare il *bound* inferiore per il cut-off, si utilizza il seguente lemma ‘tecnico’:

Lemma 2.2. *Sia $f : \Omega \rightarrow \Lambda$, con Λ insieme finito, e siano μ, ν due misure di probabilità su Ω tali che $\mathbb{E}^\nu[f] - \mathbb{E}^\mu[f] \geq r\sigma_*$ per qualche $r > 0$, dove $\mathbb{E}^\mu[f] := \sum_{x \in \Omega} \mu(x)f(x)$ e $\sigma_* := \sqrt{\max\{\text{Var}_\mu(f), \text{Var}_\nu(f)\}}$. Allora*

$$\|\mu - \nu\|_{TV} \geq 1 - \frac{8}{r^2}. \quad (19)$$

Applicando il lemma con $f(x) := \sum_j x(j)$, la somma delle coordinate dell'argomento, $\mu = \pi$, la misura stazionaria, e $\nu = \delta_{\bar{1}}$, la misura che assegna probabilità 1 al vertice $\bar{1} = (1, 1, \dots, 1)$, si ottiene che

$$d\left(\frac{1}{2}n \log n - \alpha n\right) \geq 1 - 8e^{-2\alpha+1}. \quad (20)$$

Per il *bound* superiore ricorriamo ad un processo che potrebbe sembrare scollegato dalla passeggiata aleatoria. Immaginiamo di avere due urne in cui sono distribuite n palline. Ad ogni passo scegliamo uniformemente una delle palline e le cambiamo urna. Il processo W_t su $\Omega_W := \{0, 1, \dots, n\}$ che tiene conto delle palline nella prima urna è detto processo dell'**urna di Ehrenfest**. In realtà W_t non è che una proiezione della passeggiata sull'ipercubo: basta infatti considerare W_t come $\sum_j X_t(j)$, la somma delle coordinate della passeggiata al tempo t . Vale però il sorprendente risultato:

$$\|\mathbb{P}_{\bar{1}}(X_t \in \cdot) - \pi\|_{TV} = \|\mathbb{P}_n(W_t \in \cdot) - \pi_W\|_{TV}, \quad (21)$$

dove $\mathbb{P}_{\bar{1}}$ dice che la passeggiata parte dal vertice $\bar{1}$, mentre \mathbb{P}_n indica che partiamo dalla configurazione con tutte le palline nella prima urna. Inoltre, data la simmetria dell'ipercubo, vale

$$d(t) = \|\mathbb{P}_{\bar{1}}(X_t \in \cdot) - \pi\|_{TV},$$

così che è sufficiente trovare un *bound* per la quantità di destra della (21). Ma ciò può essere facilmente fatto tramite il seguente coupling: si fanno partire due processi dell'urna di Ehrenfest e ad ogni passo si tira una moneta equa; se esce croce si fa muovere il primo processo, altrimenti il secondo. In questo modo ciascuna delle due catene sarà una versione pigra del processo. Inoltre, una volta che le due si incontrano, le si fa evolvere insieme per sempre. A questo punto, svolgendo i calcoli, si arriva a

$$\mathbb{P}\left(\tau > \frac{1}{2}n \log n + \alpha n\right) \leq \frac{c}{\sqrt{\alpha}}, \quad (22)$$

ove τ indica il primo tempo in cui le due catene si incontrano e c è una costante indipendente da n . Dal Corollario 1.3 e da (21) segue dunque che per la passeggiata aleatoria sull'ipercubo n -dimensionale si ha

$$d\left(\frac{1}{2}n \log n + \alpha n\right) \leq \frac{c}{\sqrt{\alpha}}. \quad (23)$$

3 Processi di nascita e morte

Una **catena di nascita e morte** su $\Omega_n := \{0, 1, \dots, n\}$ è una catena di Markov P tale che $P(x, y) = 0$ a meno che $|x - y| \leq 1$. Scriviamo le probabilità di transizione come

$$b(x) := P(x, x + 1) \quad \text{per } x = 0, 1, \dots, n - 1, \quad (24)$$

$$r(x) := P(x, x) \quad \forall x \in \Omega, \quad (25)$$

$$d(x) := P(x, x - 1) \quad \text{per } x = 1, 2, \dots, n \quad (26)$$

e fissiamo $d(0) = 0$, $b(n) = 0$. Se $b(0) = 0$ (oppure $d(n) = 0$) si dice che 0 (rispettivamente n) è parete assorbente (poiché 'intrappola' il processo). Se $b(0) > 0$ (rispettivamente $d(n) > 0$), è detta parete respingente.

In questo contesto siamo interessati a catene irriducibili, per cui assumeremo $b(x) > 0$ per $x \in [0, n - 1]$ e $d(x) > 0$ per $x \in [1, n]$. La distribuzione stazionaria per questo tipo di processi è

$$\pi(x) := \frac{1}{Z} \prod_{y=1}^x \frac{b(y-1)}{d(y)}, \quad (27)$$

dove Z è una costante di normalizzazione. Dato $0 < \varepsilon < 1$ definiamo il **quantile- ε** della catena come

$$Q(\varepsilon) := \min \left\{ k \in \Omega_n : \sum_{j=0}^k \pi(j) \geq \varepsilon \right\}. \quad (28)$$

Notiamo che l'irriducibilità garantisce che λ_2 , il secondo autovalore più grande della matrice di transizione, è strettamente minore di 1. Inoltre, se ammettiamo per almeno uno stato $x \in \Omega_n$ che $r(x) > 0$, evitiamo il problema della periodicità (e -1 non sarà un autovalore).

Il teorema fondamentale di questo capitolo è il seguente:

Teorema 3.1. *Per ogni $0 < \varepsilon < \frac{1}{2}$ esiste una costante esplicita $c_\varepsilon > 0$ tale che ogni catena di nascita e morte pigra ed irriducibile soddisfa*

$$t_{mix}(\varepsilon) - t_{mix}(1 - \varepsilon) \leq c_\varepsilon \sqrt{t_{rel} \cdot t_{mix}}, \quad (29)$$

dove $t_{rel} := \frac{1}{\text{gap}}$ è il tempo di rilassamento.

Nel secondo capitolo avevamo affermato che (18) è una condizione necessaria per l'esistenza del cut-off per tutte le catene di Markov. Il Teorema 3.1 implica direttamente che questa è una condizione anche necessaria per le catene di tipo nascita e morte:

Corollario 3.2. *Sia $(X_t^{(n)})$ sequenza di catene di nascita e morte pigre e irriducibili. La sequenza esibisce un cut-off se e solo se*

$$t_{mix}^{(n)} \cdot \text{gap}^{(n)} \xrightarrow{n \rightarrow \infty} \infty.$$

E' bene notare che in generale si possono costruire esempi per cui la condizione *non* è sufficiente. Un problema aperto rimane quello di caratterizzare tutte le classi di catene per cui vale il 'se e solo se'.

La dimostrazione del teorema principale si basa su stime dall'alto e dal basso di $t_{mix}(\varepsilon)$ e $t_{mix}(1 - \varepsilon)$. La prima e fondamentale osservazione è che l'ordine del tempo di mixing è quello del tempo necessario per andare dallo 0 fino al quantile- $(1 - \varepsilon)$.

Proposizione 3.3. *Se (X_t) è una catena di nascita e morte pigra e irriducibile su Ω_n , allora, per ogni $0 < \varepsilon < \frac{1}{16}$,*

$$t_{mix} \leq 16 \max \{ \mathbb{E}_0 [\tau_{Q(1-\varepsilon)}], \mathbb{E}_n [\tau_{Q(\varepsilon)}] \}, \quad (30)$$

dove per $\mathbb{E}_x [\tau_y]$ si intende il valore atteso, partendo da x , della variabile aleatoria τ_y che indica il primo tempo in cui si è raggiunto lo stato y .

Per stimare con la diseguaglianza di Chebychev $\tau_{Q(1-\varepsilon)}$ si deve poi studiare la sua varianza.

Proposizione 3.4. *Sia (X_t) una catena di nascita e morte pigra e irriducibile su Ω_n e sia gap il suo gap spettrale; allora, per $0 < \varepsilon < 1$,*

$$\text{Var}_0(\tau_{Q(1-\varepsilon)}) \leq \frac{\mathbb{E}_0 [\tau_{Q(1-\varepsilon)}]}{\varepsilon \cdot \text{gap}}. \quad (31)$$

La dimostrazione di questa proposizione segue da un teorema dimostrato da Fill in [11]. Questo risultato asserisce che, considerata una catena di nascita e morte su $\Omega_m := \{0, 1, \dots, m\}$, con m stato assorbente e tutte le altre probabilità di transizione strettamente positive ($b(x), d(x) > 0$ per ogni $x \in \Omega_m$, tranne che ovviamente $d(0) = 0$), allora, se P ha autovalori non negativi, τ_m è distribuita come la somma di m variabili aleatorie geometriche indipendenti con gli autovalori (diversi da 1) stessi come probabilità di fallimento.

Da ciò segue che

$$\text{Var}_0(\tau_m) \leq \frac{\mathbb{E}_0[\tau_m]}{\text{gap}_m}. \quad (32)$$

Prendendo ora $m = Q(1 - \varepsilon)$ e confrontando i gap spettrali delle catene su Ω_n e $\Omega_{Q(1-\varepsilon)}$ (risulta, con ovvie notazioni, $\text{gap}_m \geq \varepsilon \cdot \text{gap}_{Q(1-\varepsilon)}$), si ottiene finalmente (31).

A questo punto il teorema principale è piuttosto agilmente dimostrabile se t_{rel} non è molto più piccolo di t_{mix} . In caso contrario, per completare le stime, si ha bisogno di un lemma tecnico.

Lemma 3.5. *Sia (X_t) catena di nascita e morte pigra ed irriducibile su Ω_n e supponiamo che esista $\varepsilon \in (0, \frac{1}{16})$ per cui*

$$t_{rel} < \varepsilon^4 \mathbb{E}_0[\tau_{Q(1-\varepsilon)}]. \quad (33)$$

Allora, per ogni $\varepsilon \leq \alpha \leq \beta < 1 - \varepsilon$,

$$\mathbb{E}_{Q(\alpha)}[\tau_{Q(\beta)}] \leq \frac{3}{2\varepsilon} \sqrt{t_{rel} \cdot \mathbb{E}_0[\tau_{Q(\frac{1}{2})}]}. \quad (34)$$

4 La dinamica di Glauber per il modello di Ising

Consideriamo lo spazio degli stati $\Omega := \{-1, +1\}^V$, dove V è l'insieme dei vertici di un grafo completo. Da un punto di vista fisico possiamo immaginare in ogni vertice del grafo un piccolo magnete che può puntare all'insù (se gli abbiamo assegnato il valore, o **spin**, $+1$) o all'ingiù (se lo spin vale -1); tutti i magneti si influenzano tra loro, provando a puntare nella stessa direzione. Questa interazione è regolata da un parametro $\beta \geq 0$, che può essere interpretato come l'inverso della temperatura dell'ambiente. Più la temperatura è bassa, più l'interazione tra spin è forte.

Nel **modello di Ising** si assegna ad ogni possibile configurazione di spin $\sigma \in \Omega$ un'energia

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

dove $\sigma(v)$ indica lo spin della configurazione σ nel vertice v . Come si può vedere, l'energia decresce se il numero di coppie di siti con lo stesso spin

aumenta. Infine assegnamo ad ogni configurazione una probabilità (detta **misura di Gibbs**):

$$\mu(\sigma) := \frac{e^{-\beta H(\sigma)}}{Z(\beta)}, \quad (35)$$

ove $Z(\beta)$ è la cosiddetta funzione di partizione, con funzione normalizzante.

Il processo che vogliamo considerare è la **dinamica di Glauber** di tipo *heat-bath*: ad ogni passo selezioniamo uniformemente un vertice $v \in V$, "cancelliamo" lo spin che vi si trova e lo rimpiazziamo con un "+" con probabilità $p^+(S(\sigma) - \frac{\sigma(v)}{n})$, dove $S(\sigma) := \frac{1}{n} \sum_{v \in V} \sigma(v)$ è la magnetizzazione di σ , e

$$p^+(x) := \frac{e^{\beta x}}{e^{\beta x} + e^{-\beta x}} = \frac{1 + \tanh(\beta x)}{2}; \quad (36)$$

altrimenti lo spin viene rimpiazzato con un "-". In questo modo la dinamica risulta essere reversibile rispetto a μ . Visto che le funzioni p^+ e p^- sono sempre strettamente positive, la catena è aperiodica, e siccome è possibile andare da una qualsiasi configurazione ad ogni altra configurazione in al più n passi, la catena è anche irriducibile. Denoteremo la dinamica di Glauber con (X_t) .

Il **grand coupling** è una tecnica che prevede che da ogni possibile configurazione di Ω parta una copia della dinamica. Tutte queste catene sono dipendenti da una stessa fonte aleatoria: ad ogni passo si estrae un vertice w a caso tra gli n , una variabile $U \sim Unif([0, 1])$ e per ogni dinamica X_t^σ partente da qualche σ si pone uno spin "+" in w se $p_+(S(X_t^\sigma) - X_t^\sigma(e)/n) \geq U$ e un "-" altrimenti. Questa tecnica è anche detta del **coupling monotono**.

Una catena associata ad (X_t) di fondamentale importanza è la **catena della magnetizzazione**, definita come $S_t := S(X_t) = \frac{1}{n} \sum_{v=1}^n X_t(v)$, con spazio degli stati $\Omega_S := \{-1, -1 + \frac{2}{n}, \dots, 1 - \frac{2}{n}, 1\}$.

Infine chiamiamo la **catena a due stati** la seguente proiezione della (X_t) : fissiamo $\sigma_0 \in \Omega_0 := \{\sigma \in \Omega : |S(\sigma)| \leq \frac{1}{2}\}$ e chiamiamo $\bar{u}_0 := |\{i : \sigma_0(i) = +1\}|$ e $\bar{v}_0 := |\{i : \sigma_0(i) = -1\}|$. Per $\sigma \in \Omega$ siano

$$U(\sigma) = U_{\sigma_0}(\sigma) := |\{i : \sigma(i) = \sigma_0(i) = +1\}|$$

$$V(\sigma) = V_{\sigma_0}(\sigma) := |\{i : \sigma(i) = \sigma_0(i) = -1\}|.$$

Allora $(U_t, V_t)_{t \geq 0} := (U(X_t), V(X_t))_{t \geq 0}$ è una catena di Markov con probabilità di transizioni che dipendono da σ_0 sullo spazio degli stati $\{0, \dots, \bar{u}_0\} \times \{0, \dots, \bar{v}_0\}$ e con una sua misura stazionaria π_2 .

Enunciamo alcuni risultati riguardanti le catene 'secondarie' che risulteranno fondamentali per la dimostrazione del cut-off:

Lemma 4.1. *Sia (X_t, \tilde{X}_t) coupling monotono partente da σ e $\tilde{\sigma}$. Definiamo $\rho := 1 - \frac{1}{n} - \tanh(\frac{\beta}{n})$. Siano $S_t = S(X_t^\sigma)$ e $\tilde{S}_t = S(X_t^{\tilde{\sigma}})$. Allora*

$$\mathbb{E}_{\sigma, \tilde{\sigma}}[|S_t - \tilde{S}_t|] \leq \frac{2}{n} \rho^t \text{dist}(\sigma, \tilde{\sigma}),$$

dove $\text{dist}(\eta, \eta') = \frac{1}{2} \sum_i |\eta(i) - \eta'(i)|$.

Lemma 4.2. *Per la catena della magnetizzazione $S(t)$, $\forall s, \tilde{s} \in \Omega_S$ si ha che*

$$|\mathbb{E}_s[S_1] - \mathbb{E}_{\tilde{s}}[S_1]| \leq |s - \tilde{s}|.$$

Teorema 4.3. *Siano $\sigma, \tilde{\sigma} \in \Omega$ tali che $S(\sigma) = S(\tilde{\sigma})$, $R_0 := U(\tilde{\sigma}) - U(\sigma) > 0$ e $R_t := U(\tilde{X}_t) - U(X_t)$. Sia inoltre*

$$\Xi := \left\{ \sigma \in \Omega : \min\{U(\sigma), \bar{u}_0 - U(\sigma), V(\sigma), \bar{v}_0 - V(\sigma)\} \geq \frac{n}{16} \right\}.$$

\exists un coupling (X_t, \tilde{X}_t) della dinamica di Glauber con $X_0 = \sigma$, $\tilde{X}_0 = \tilde{\sigma}$ tale che

- (i) $S(X_t) = S(\tilde{X}_t) \forall t \geq 0$;
- (ii) $R_t \geq 0 \forall t \geq 0$ e $\mathbb{E}_{\sigma, \tilde{\sigma}}[R_{t+1} - R_t | X_t, \tilde{X}_t] \leq 0$;
- (iii) $\exists c > 0$ non dipendente da n tale che se $X_t \in \Xi$ e $\tilde{X}_t \in \Xi$, allora $\mathbb{P}_{\sigma, \tilde{\sigma}}(R_{t+1} - R_t \neq 0 | X_t, \tilde{X}_t) \geq c$.

Descriviamo brevemente il coupling del precedente lemma: per X_t si sceglie in maniera casuale uno degli n vertici, sia v , e vi si pone normalmente uno spin secondo i valori di p_+ e di p_- . Per \tilde{X}_t si sceglie invece un vertice che abbia lo *stesso* spin di $X_t(v)$ e lo si aggiorna con lo *stesso* spin con cui si è aggiornata la prima catena (notiamo che anche \tilde{X}_t segue la dinamica di Glauber in quanto le due catene hanno, per ipotesi, e continuano ad ogni passo ad avere, per costruzione, la stessa magnetizzazione).

Ecco il teorema che descrive il cut-off per la dinamica di Glauber:

Teorema 4.4. *Sia (X_t) la dinamica di Glauber per il modello di Ising sull' n -grafo completo. Sia μ la sua misura stazionaria. Se $\beta < 1$, si ha cut-off al tempo $t_n := \frac{n \log n}{2(1-\beta)}$ con finestra di ordine n .*

La dimostrazione del bound superiore è la più laboriosa. Dividiamo la sua dimostrazione in tre diverse fasi.

Nella FASE1 facciamo procedere la catena per $t_0 := \theta_0 n$ passi affinché ci si trovi con alta probabilità in una “buona” posizione (cioè in $\Omega_0 := \{\sigma \in \Omega : |S(\sigma)| \leq \frac{1}{2}\}$). A questo punto si fisserà *una volta per tutte* $\sigma_0 \in \Omega_0$. Questa scelta si rivelerà fondamentale per un corretto svolgimento della FASE3.

Nella FASE2 utilizziamo un coupling che faccia partire da σ_0 e da un qualsiasi altro $\tilde{\sigma} \in \Omega$ due copie della catena, con l’obiettivo di far coincidere le loro magnetizzazioni con alta probabilità dopo $O\left(\frac{n \log n}{2(1-\beta)}\right)$ passi. La regola d’evoluzione del coupling varia però nel tempo: per t_n passi facciamo evolvere le due copie della catena secondo la regola del coupling monotono.

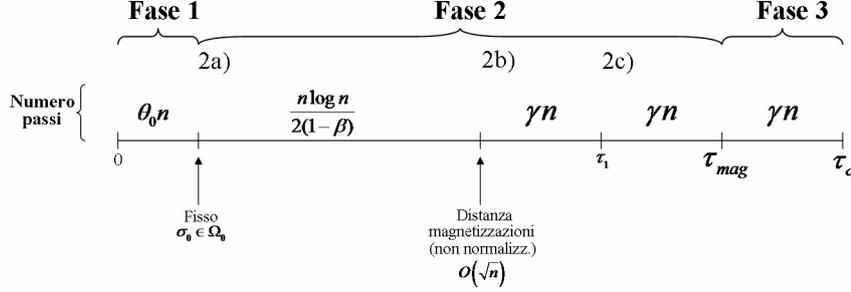


Figure 2: *Time line per la dimostrazione del bound superiore. Si consideri che gli eventi segnati non sono certi, ma avvengono con alta probabilità.*

Otteniamo così, dal Lemma 4.1, $\mathbb{E}_{\sigma, \tilde{\sigma}} \left[\frac{n}{2} |S_{t_n} - \tilde{S}_{t_n}| \right] \leq \text{cost} \sqrt{n}$. Sia poi $\tau_1 := \min \{t \geq t_n : \frac{n}{2} |S_t - \tilde{S}_t| \leq 1\}$. Fino a τ_1 facciamo evolvere indipendentemente le due catene. Per il Lemma 4.2 si ha che $(S_t - \tilde{S}_t)_{t_n \leq t \leq \tau_1}$ ha *drift* negativo; in particolare è possibile far vedere che $\mathbb{P}_{\sigma, \tilde{\sigma}}(\tau_1 > t_n + 2\gamma n) \leq O(\gamma^{-\frac{1}{2}})$. A questo punto, nel caso in cui in τ_1 le magnetizzazioni non siano ancora uguali, basteranno altri $O(n)$ passi per farle coincidere, grazie ad una versione leggermente modificata del coupling monotono. In conclusione, nella FASE2, otteniamo che le magnetizzazioni delle due catene si incontreranno entro $t_n + 3\gamma n$ con probabilità almeno $1 - O(\gamma^{-\frac{1}{2}})$.

Per la FASE 3 si osserva che

$$\max_{\sigma_0 \in \Omega_0} \|\mathbb{P}_{\sigma_0}(X_t \in \cdot) - \mu\|_{TV} = \max_{(\bar{u}_0, \bar{v}_0) \in \Lambda_0} \|\mathbb{P}_{(\bar{u}_0, \bar{v}_0)}((U_t, V_t) \in \cdot) - \pi_2\|_{TV}. \quad (37)$$

Vorremmo a questo punto utilizzare Theorem 4.3 con $\sigma = X_{t_n+3\gamma n}$, $\tilde{\sigma} = \tilde{X}_{t_n+3\gamma n}$ e con il σ_0 trovato in FASE 1 come riferimento. Sotto l'evento {le magnetizzazioni si siano già incontrate nella FASE2}, si può dimostrare che valgono i punti (i), (ii); inoltre (iii) sussiste con probabilità $1 - O(n^{-1})$. Il teorema garantisce allora che le due catene a due coordinate si incontreranno con alta probabilità dopo altri γn passi.

Ricordando allora (37), richiamando il Corollario 1.3 e facendo tendere n e γ all'infinito otteniamo il *bound* cercato.

Per la trovare una stima inferiore, invece, è sufficiente notare che la catena della magnetizzazione è una proiezione della dinamica di Glauber. Basta pertanto trovare un bound per la distanza di (S_t) dalla sua misura stazionaria.

Per far ciò è sufficiente studiare un poco più in dettaglio le probabilità di transizione di (S_t) e ottenere, applicando Chebychev,

$$\begin{aligned} \|\mathbb{P}_{s_0} \left(S_{(t_n - \frac{\alpha n}{(1-\beta)})} \in \cdot \right) - \pi_S\|_{TV} &\geq \pi_S(D) - \mathbb{P}_{s_0} \left(|S_{(t_n - \frac{\alpha n}{(1-\beta)})}| \in D \right) \\ &\geq 1 - \frac{c}{s_0^2 e^{2\alpha}} \end{aligned} \quad (38)$$

per appropriati $s_0 \in \Omega_S$ e $D \subset \Omega_S$.

Per il valore critico $\beta = 1$ è possibile invece dimostrare che l'ordine del tempo di mixing aumenta fino ad $O(n^{\frac{3}{2}})$ e che per temperature ancora minori ($\beta > 1$) il tempo di mixing diventa addirittura esponenziale. Inoltre in nessuno dei due casi si verifica il fenomeno del cut-off.

5 Stime analitiche

Data una catena di Markov a tempo discreto P , è sempre possibile associarle un processo $(X_t)_{t \in \mathbb{R}^+}$ a tempo continuo nella seguente maniera: invece di effettuare ad ogni unità di tempo un passo secondo la matrice di transizione, si attende tra un passo e l'altro un tempo distribuito come una variabile aleatoria esponenziale di parametro 1.

Ponendo $Q = P - Id$, si ha che la probabilità di trovarsi nel sito y dopo un tempo t essendo partiti da x è uguale a

$$e^{tQ}(x, y) = \sum_{k=0}^{\infty} \frac{e^{-t} t^k}{k!} P^k(x, y) =: H_t(x, y). \quad (39)$$

La matrice H_t è detta **heat-kernel** del processo.

E' semplice verificare che se π è misura stazionaria per P , allora lo è anche per (X_t) . Inoltre, uno dei vantaggi delle catene a tempo continuo è che esse evitano il problema della periodicità. Il Teorema Ergodico 1.1 ha il seguente corrispettivo nel caso continuo:

Teorema 5.1. *Sia P matrice stocastica irriducibile e sia H_t l'heat-kernel corrispondente. Esiste un'unica distribuzione di probabilità π tale che, per ogni $t \geq 0$,*

$$\pi H_t = \pi.$$

Inoltre

$$\max_{x \in \Omega} \|H_t(x, \cdot) - \pi\|_{TV} \xrightarrow{t \rightarrow \infty} 0.$$

Data una matrice di transizione P ed la catena di Markov a tempo continuo (X_t) ad essa associato, possiamo definire la forma di Dirichlet del processo come

$$\mathcal{D}(f, f) = \sum_{x, y \in \Omega} (f(y) - f(x))^2 \pi(x) Q(x, y).$$

E' risaputo che il gap spettrale γ di P verifica

$$\gamma = 1 - \lambda_2 = \min_{\substack{f \in \mathbb{R}^\Omega \\ \text{Var}_\pi(f) \neq 0}} \left\{ \frac{\mathcal{D}(f, f)}{\text{Var}_\pi(f)} \right\}.$$

La stima superiore in (11) ha il suo equivalente continuo:

$$t_{mix}^{cont} \left(\frac{1}{2e} \right) \leq \frac{1}{2\gamma} \left(2 + \log \left(\frac{1}{\pi_{min}} \right) \right), \quad (40)$$

dove $t_{mix}^{cont}(\cdot)$ è il corrispettivo continuo del tempo di mixing e $\pi_{min} = \min_{x \in \Omega} \pi(x)$.

La **costante logaritmica di Sobolev** è definita in maniera analoga alla rappresentazione variazionale del gap spettrale, in cui la varianza di f viene sostituita dell'**entropia** di f^2 :

$$\text{Ent}_\pi(f) = \mathbb{E}_\pi [f \log(f)] \quad \text{per } f \geq 0, \pi(f) = 1. \quad (41)$$

Quindi

$$\alpha := \inf_{\substack{f \in \mathbb{R}^\Omega \\ \pi(f^2) = 1}} \left\{ \frac{\mathcal{D}(f, f)}{\text{Ent}_\pi(f^2)} \right\}. \quad (42)$$

Lemma 5.2. *Sia P catena di Markov reversibile e irriducibile; allora, per la catena a tempo continuo ad essa associata si ha*

$$\frac{1}{2\alpha} \leq t_{mix}^{cont} \left(\frac{1}{2e} \right) \leq \frac{1}{4\alpha} \left(4 + \log_+ \log \frac{1}{\pi_{min}} \right). \quad (43)$$

Per rendere la dinamica di Glauber a tempo continuo, possiamo immaginare di avere una sveglia che attenda un tempo distribuito come una variabile esponenziale di parametro n prima di suonare. Ad ogni suo trillo, viene effettuata una mossa della dinamica a tempo discreto. La forma di Dirichlet del processo è, per ogni $f \in \mathbb{R}^\Omega$,

$$\mathcal{D}(f, f) := \sum_{\substack{\sigma \in \Omega \\ i=1,2,\dots,n}} (f(\sigma^i) - f(\sigma))^2 \mathcal{L}(\sigma, \sigma^i) \pi(\sigma), \quad (44)$$

dove

$$\mathcal{L}f(\sigma) = \sum_i \sum_{x=\pm 1} \mathbb{P}(\sigma_i = x | \{\sigma_j\}_{j \neq i}) (f(\sigma^{i,x}) - f(\sigma)). \quad (45)$$

La catena della magnetizzazione (questa volta non normalizzata: $\Omega_S := \{-n, -n+2, \dots, n-2, n\}$) corrispondente ha misura stazionaria

$$\mu(s) := \sum_{\sigma: S(\sigma)=s} \mu(\sigma).$$

Utilizzando la formula di approssimazione di Stirling si vede che $\mu(s)$ ha la forma di una gaussiana di media 0 e varianza $\sim \frac{2n}{1+\beta}$. Grazie alle formule di Miclo (vedi, per es. [3]) si può mostrare che il gap della catena della magnetizzazione (che indicheremo con gap') soddisfa

$$\frac{1}{\text{gap}'} \leq cn, \quad (46)$$

dove c è una costante indipendente dalla taglia del grafo n .

Studiamo il gap spettrale della dinamica di Glauber. Per farlo, notiamo che, dalla formula della ‘varianza totale’, possiamo riscrivere la varianza di una qualsiasi funzione f come

$$\text{Var}(f) = \mathbb{E}[\text{Var}(f|s)] + \text{Var}(\mathbb{E}[f|s]). \quad (47)$$

Per stimare il primo addendo della (47) osserviamo che, fissando il valore della magnetizzazione, ogni configurazione del nostro modello di Ising può essere vista come una configurazione del modello di Bernoulli-Laplace (vedi la versione estesa della tesi o, per es., [8] per una descrizione del modello), interpretando i "+" come particelle ed i "-" come spazi vuoti. Ma per il modello di Bernoulli-Laplace è ben noto che il gap spettrale (indipendentemente dal numero dal numero di siti e di particelle coinvolte) è semplicemente una costante, e quindi

$$\text{Var}(f|s) = \text{Var}^{BL}(f) \leq \frac{1}{(\text{gap}^{BL})} \mathcal{D}^{BL}(f, f) = c \cdot \mathcal{D}^{BL}(f, f), \quad (48)$$

dove gli apici ‘BL’ si riferiscono al modello di Bernoulli Laplace e $c > 0$ è una costante. Mediando sulle possibili magnetizzazioni e riscrivendo la forma di Dirichlet di Bernoulli-Laplace come la forma di Dirichlet della dinamica di Glauber (a meno di una costante) si ottiene infine

$$\mathbb{E}[\text{Var}(f|s)] \leq \text{cost.} \cdot \mathcal{D}(f, f). \quad (49)$$

Per la seconda parte di (47), notiamo che $g(s) := \mathbb{E}[f|s]$ è una funzione della sola magnetizzazione. Perciò, dalla (46),

$$\text{Var}(g(s)) \leq \frac{1}{\text{gap}'} \mathcal{D}^{mag}(g, g) \leq cn \mathcal{D}^{mag}(g, g) \leq \text{cost.} \cdot \mathcal{D}(f, f), \quad (50)$$

dove per l’ultima disuguaglianza abbiamo dovuto riscrivere $\mathcal{D}^{mag}(g, g)$, la forma di Dirichlet della catena della magnetizzazione applicata a g , come $\frac{1}{n} \mathcal{D}(f, f)$ (a meno di una costante positiva). Inserendo (49) e (50) in (47), otteniamo che il gap spettrale della dinamica di Glauber per il modello di Ising su grafo completo è inferiormente limitato da una costante strettamente positiva indipendente dal numero di vertici del grafo.

Sarebbe auspicabile poter trovare un *bound* inferiore costante anche per la costante logaritmica di Sobolev della dinamica di Glauber. Anche in questo caso la strategia è la stessa: riscriviamo l'entropia di una qualsiasi funzione, condizionandola al valore della magnetizzazione, come

$$\text{Ent}_\mu(f) = \mathbb{E} \left[\mathbb{E}_\mu [f|s] \cdot \text{Ent}_\mu \left(\frac{f}{\mathbb{E}_\mu [f|s]} \middle| s \right) \right] + \text{Ent}_\pi(\mathbb{E}_\mu [f|s]). \quad (51)$$

Per la $\text{Ent}_\pi(\mathbb{E}_\mu [f|s])$ è possibile seguire completamente il ragionamento fatto per il gap spettrale. Infatti, anche per la log-Sobolev della catena della magnetizzazione è possibile vedere con le formule di Miclo che

$$\frac{1}{\alpha_S} \leq cn. \quad (52)$$

I problemi si presentano invece nella stima del primo addendo. La log-Sobolev del modello di Bernoulli-Laplace nel peggiore dei casi è infatti stimabile solo con

$$\frac{1}{\alpha_{n,r}^{BL}} \leq c \log n, \quad (53)$$

come dimostrato nell'articolo [15]. Seguendo ancora una volta il ragionamento fatto nel caso del gap siamo perciò in grado di affermare solamente che

$$\mathbb{E} \left[\mathbb{E}_\mu [f|s] \cdot \text{Ent}_\mu \left(\frac{f}{\mathbb{E}_\mu [f|s]} \middle| s \right) \right] \leq \text{cost} \cdot \log(n) \cdot \mathcal{D}(f, f). \quad (54)$$

Teorema 5.3. *Per $\beta < 1$, esiste una costante $c > 0$ tale che la costante log-Sobolev della dinamica di Glauber per il modello di Ising sul grafo n -completo verifica*

$$\alpha \geq c \cdot \frac{1}{\log n}. \quad (55)$$

E' probabile, però, che questo risultato sia migliorabile.

References

- [1] Aldous D. and Diaconis P. (1986) *Shuffling cards and stopping times*, Amer. Math. Monthly **93**, 333-348.
- [2] 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>.
- [3] Bodineau T. and Martinelli F. (2002) *Some new results on the kinetic Ising model in a pure phase*, arXiv:math-ph/0202013v1.
- [4] Casella G. and Robert C. (2008) *A History of Markov Chain Monte Carlo*, <http://www.stat.ufl.edu/~casella/Papers/MCMCHistory.pdf>.
- [5] 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>.
- [6] Diaconis P. (2008) *The Markov Chain Monte Carlo Revolution*, Bull. Amer. Math. Soc. Nov. 2008, <http://www-stat.stanford.edu/~cgates/PERSI/papers/MCMCRev.pdf>.
- [7] Diaconis P. and Saloff-Coste L. (2006) *Separation cut-offs for birth and death chains*, Ann. Appl. Probab. **16**, no.4, 2098-2122.
- [8] Diaconis P. and Shahshahani M. (1987) *Time to reach stationarity in the Bernoulli-Laplace diffusion model*, SIAM J. Math. Anal. **18**, 208-218.
- [9] Ding J., Lubetzky E. and Peres Y. (2008) *The mixing time evolution of Glauber dynamics for the mean-field Ising model*, arXiv:0806.1906v2.
- [10] Ding J., Lubetzky E. and Peres Y. (2008) *Total variation cutoff in birth and death chains*, arXiv:0801.2625v4.
- [11] Fill J. A. (2009) *The passage time distribution for a birth-and-death chain: Strong stationary duality gives a first stochastic proof*, Journal of Theoretical Probability, to appear.
- [12] Gao F. and Quastel J. (2003) *Exponential decay of entropy in the random transposition and Bernoulli-Laplace models*, Ann. Appl. Probab. **13**, no.4, 1591-1600.
- [13] Griffiths R. B., Weng C.-Y. and Langer J. S. (1966) *Relaxation Times for Metastable States in the Mean-Field Model of a Ferromagnet*, Phys. Rev. **149**, 301-305.
- [14] Karlin S. and McGregor J. (1959) *Coincidence properties of birth and death processes*, Pacific J. Math. **9**, 1109 -1140.

- [15] Lee T. and Yau H. (1998) *Logarithmic Sobolev inequality for some models of random walks*, Annals of Probability, **26** , no.4, 1855-1873.
- [16] 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.
- [17] Levin D. A., Peres Y. and Wilmer E. L. (2009) *Markov Chains and Mixing Times*, American Mathematical Society.
- [18] Mitzenmacher M., Upfal E. (2005) *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*, Cambridge University Press, Cambridge.
- [19] Norris J. R. (1998) *Markov chains*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 2, Cambridge University Press, Cambridge. Reprint of 1997 original.
- [20] Reed M. and Simon B. (1978) *Analysis of Operators*, Methods of modern mathematical physics, Vol. 4, Academic Press, New York.
- [21] Saloff-Coste L. (1997) *Lectures on finite Markov chains*, Lectures On Probability Theory and Statistics, Ecole d'Ete de Probabilites de Saint-Flour XXVI, 301-413.
- [22] Shiryaev A. N. (1980) *Probability*, Graduate Texts in Mathematics, Springer.
- [23] Williams D. (1991) *Probability with martingales*, Cambridge Mathematical Textbooks, Cambridge University press, Cambridge.