Markov Chain Monte Carlo

Enrico Scalas

Bilbao, 16-17 April 2013

Markov Chain Monte Carlo

Outline

- Discrete Markov chains
- Example: The Ehrenfest Urn model
- The algorithm of Metropolis
- Example: The Ising model

Markov Chain Monte Carlo

Discrete homogenous Markov chains

Definition (Discrete Homogeneous Markov chain)

Let $X_1, \ldots, X_n, \ldots$ be a sequence of random variables with each X_i taking values in a finite or countable state space $\mathcal{S}$. For $k = 0, 1, 2, \ldots$, define the k-step transition probabilities as

$$p_{i,j}^{(k)} = \text{Pr}(X_{t+k} = j | X_t = i) \quad (i, j \in \mathcal{S}). \tag{1}$$

We usually write $p_{i,j} = p_{i,j}^{(1)}$. The Markov chain is *homogeneous* meaning that the *transition probability matrix* $P = (p_{i,j})$ does not depend on t; one further has that $p_{i,j}^{(k)}$ is the (i, j)-th entry of P^k.

Markov Chain Monte Carlo

Discrete homogenous Markov chains

Definition (Irreducibility)

A Markov chain is called *irreducible*, if any state can be reached from any other state. In other words, for any couple $i, j \in \mathcal{S}$, there exists a k such that $p_{i,j}^{(k)} > 0$.

Remark

This property is sometimes called *ergodicity*.

Definition (Periodicity)

Let D be a positive integer. An irreducible chain has *period* D if D is the greatest common divisor of $\{k \geq 1 : p_{i,i}^{(k)} > 0\}$ for some $i \in \mathcal{S}$ (or, equivalently, for all $i \in \mathcal{S}$). A chain of period 1 is called *aperiodic*. If an irreducible chain has $p_{i,i} > 0$ for some i, then it is aperiodic.

Markov Chain Monte Carlo

Discrete homogenous Markov chains

Theorem(Long-run behaviour of discrete Markov chains) Consider an aperiodic and irreducible Markov chain with state space $\mathcal{S}$. For every $i, j \in \mathcal{S}$, the limit

$$\pi_j = \lim_{k\to\infty} p_{i,j}^{(k)} \tag{2}$$

exists and is independent of i. Moreover, one has that

1. if $\mathcal{S}$ is finite, then $\sum_{j\in\mathcal{S}} \pi_j = 1$ and $\pi_j = \sum_{i\in\mathcal{S}} \pi_i p_{i,j}$ for every $j \in \mathcal{S}$. This means that if π is the row vector with entries π_i, then $\pi = P\pi$. The only solution to the problem

$$v = vP, \text{ with } \sum_{i\in\mathcal{S}} v_i = 1 \text{ and } v_i \geq 0 \tag{3}$$

   is $v = \pi$;

2. if $\mathcal{S}$ is countable, then either $\pi_j = 0$ for every j, in which case (3) has no solution or π satisfies (3) and is the only solution of (3).

Markov Chain Monte Carlo

Discrete homogenous Markov chains

Remark

The previous theorem can be interpreted as follows

- $\pi_i \simeq \text{Pr}(X_t = i)$ for large t, independent of X_0 (or of the initial distribution).

- π_i is the long-run fraction of the time spent by the chain in state i. In other words

$$\pi_i = \lim_{t \to \infty} \frac{\sum_{k=1}^{t} \mathbf{1}(X_k = i)}{t}$$

  with probability 1. This result is particularly important for Markov chain Monte Carlo, as a single realization of the chain can be used to sample π.

- When π exists (this is also called the *positive recurrent* case), it is the *equilibrium* or *stationary* or *steady state* distribution, meaning that if X_0 has distribution π, then X_t has exactly the distribution π for every time t.

Markov Chain Monte Carlo

Discrete homogenous Markov chains

Proposition (Detailed balance)

Consider an irreducible Markov chain with discrete state space $\mathcal{S}$. Assume that there exist positive numbers π_i, with $i \in \mathcal{S}$, such that $\sum_{i \in \mathcal{S}} \pi_i = 1$ and that *detailed balance* holds, for every $i, j \in \mathcal{S}$

$$\pi_i p_{i,j} = \pi_j p_{j,i}, \tag{4}$$

then $\pi = (\pi_i)_{i \in \mathcal{S}}$ is the equilibrium distribution.

Definition (Reversibility)

If there exist positive numbers π_i such that detailed balance (equation (4)) holds, then the chain is called *reversible* (with respect to π).

Remark (Master equation)

Detailed balance can be discussed starting from the so-called *master equation*

$$\Pr(X(t+1) = j) - \Pr(X(t) = j) = \sum_{i \in \mathcal{S}} [\Pr(X(t) = i) p_{i,j} - \Pr(X(t) = j) p_{j,i}]. \tag{5}$$

Markov Chain Monte Carlo

Discrete homogenous Markov chains

Remark (Meaning of reversibility)

Consider an irreducible and aperiodic Markov chain at equilibrium and reversible with respect to π then

$$\Pr(X_0 = i, X_1 = j, X_2 = k) = \pi_i p_{i,j} p_{j,k} = \pi_j p_{j,i} p_{j,k} = \pi_k p_{k,j} p_{j,i} = \Pr(X_0 = k, X_1 = j, X_2 = i).$$

In other words, if one watches a movie of the chain, one cannot say whether it moves forwards or backwards.

Definition (Symmetry)

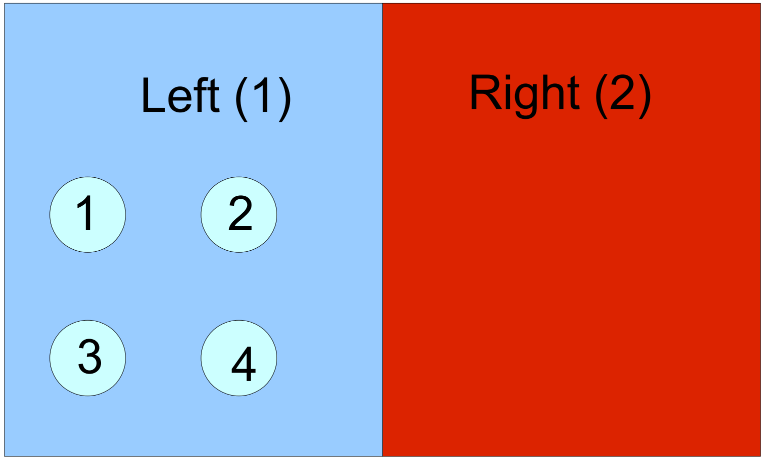
A Markov chain is *symmetric* if $p_{i,j} = p_{j,i}$ for every $i, j \in \mathcal{S}$.

Corollary (Uniform distribution) An irreducible symmetric Markov chain on the finite state space $\mathcal{S}$ has a uniform equilibrium distribution $\pi_i = 1/|\mathcal{S}|$.

Markov Chain Monte Carlo

Discrete homogenous Markov chains

Example: The Ehrenfest urn



n objects in $g = 2$ categories or classes. In this case $n = 4$ marbles and $g = 2$ boxes.

Individual descriptions:

$(X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 1)$ means that ball 1 is in class 1 (in the left urn) etc..

Statistical descriptions:

$(Y_1 = 4, Y_2 = 0)$ means that all the balls are in class 1 (in the left urn).

There are g^n possible individual descriptions and $\binom{n+g-1}{g-1}$ statistical descriptions.

A single statistical description corresponds to $n!/(n_1!n_2!\ldots n_g!)$ individual descriptions.

Markov Chain Monte Carlo

Discrete homogenous Markov chains

Example: The Ehrenfest urn

At every step, a marble is randomly selected (selected with uniform distribution) and moved in the other box. This is a periodic version of the Ehrenfest urn, with period $D = 2$. Let us enumerate the five statistical descriptions using them as space state $\mathcal{S}$, we write $(4, 0)$ for $(Y_1 = 4, Y_2 = 0)$ etc.:

$$\mathcal{S} = \{(4,0), (3,1), (2,2), (1,3), (0,4)\}.$$

Let k be the marbles in the left urn, then, with our choice of $\mathcal{S}$, for $k = \{1, 2, 3\}$, the transition probability matrix is $p_{k,k+1} = (n-k)/n$ and $p_{k,k-1} = k/n$ with all the other transitions forbidden, for $k = 0$, we move to $k = 1$ for sure and for $k = 4$, we go to $k = 3$ with probability 1:

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/4 & 0 & 3/4 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 3/4 & 0 & 1/4 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

The equilibrium distribution is

$$\pi = (1/16, 4/16, 6/16, 4/16, 1/16).$$

Markov Chain Monte Carlo

Discrete homogenous Markov chains

Example: The Ehrenfest urn

For an aperiodic version, assume that, after choosing the marble, we toss a coin. If we get head, the marble stays where it is, otherwise, we move it to the other box. With the same notation as before, for $k = \{1, 2, 3\}$, we have $p_{k,k+1} = (n-k)/2n$, $p_{k,k} = 1/2$, and $p_{k,k-1} = k/2n$; for $k = 0$, we go to $k = 1$ with probability $1/2$ or we stay in $k = 0$ with probability $1/2$. The same happens for $k = 4$. The new transition probability matrix is

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/8 & 1/2 & 3/8 & 0 & 0 \\ 0 & 1/4 & 1/2 & 1/4 & 0 \\ 0 & 0 & 3/8 & 1/2 & 1/8 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix}.$$

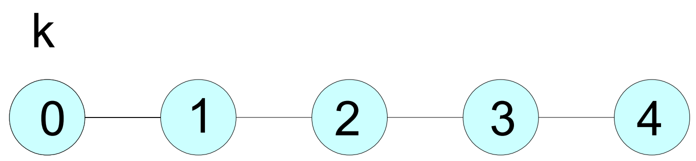
The equilibrium distribution is the same as before

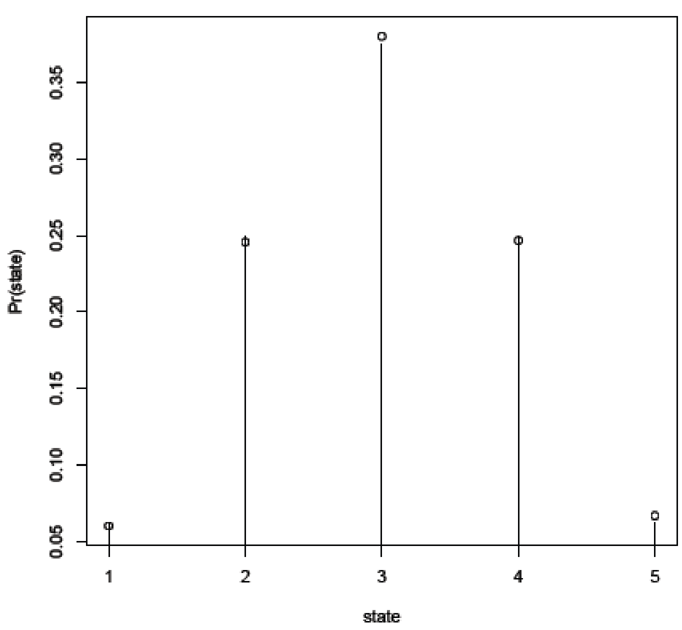
$$\pi = (1/16, 4/16, 6/16, 4/16, 1/16).$$

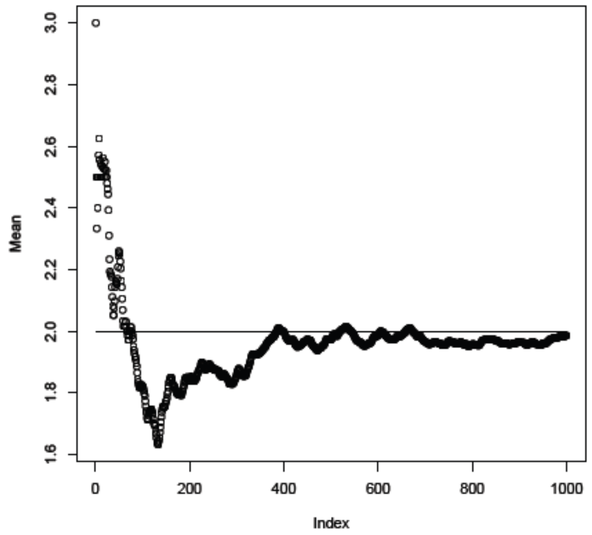
Markov Chain Monte Carlo

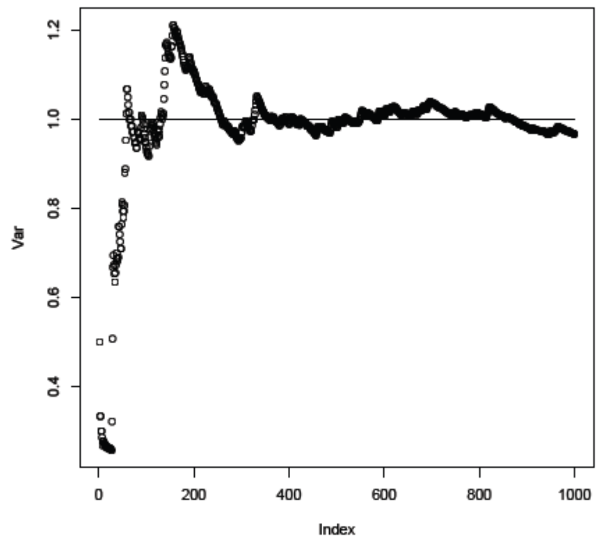
Discrete homogenous Markov chains

Example: The Ehrenfest urn, Monte Carlo simulation









Markov Chain Monte Carlo

The algorithm of Metropolis

This is the basic idea behind Markov chain Monte Carlo methods: You have a probability distribution f on a set $\mathcal{S}$ and you want to generate random elements of $\mathcal{S}$ with distribution f. You create an irreducible and aperiodic Markov chain whose equilibrium distribution is f and then you simulate the chain for many steps. In the long run, the states of the chain should follow the target distribution. The Markov chain works for you.

Markov Chain Monte Carlo

The algorithm of Metropolis

Let $\mathcal{S}$ be a finite state space. Let Q be a symmetric transition probability matrix. Let π be our target probability distribution on $\mathcal{S}$ with $\pi_i > 0$ for every $i \in \mathcal{S}$. In the Metropolis algorithm, we build a new Markov chain $X_0, \ldots, X_n, \ldots$ with the same state space. Let $X_t = i$ be the current state, the next state is determined according to the following algorithm.

1. Choose $Y = j$ according to the transition probability Q

   $$\Pr(Y = j | X = i) = q_{i,j},$$

   $Y = j$ is the *trial state*.

2. Compute $\alpha = \min\{1, \pi_Y / \pi_i\}$. This is the *acceptance probability*.

3. Accept $Y = j$ with probability α. Namely, generate $U \sim U[0, 1]$. If $U \leq \alpha$ accept the proposal and set $X_{t+1} = Y$, otherwise reject the proposal and set $X_{t+1} = X_t$.

Markov Chain Monte Carlo

The algorithm of Metropolis

The Metropolis algorithm defines a Markov chain with transition probabilities given by

$$p_{i,j} = q_{i,j} \min\{1, \pi_j/\pi_i\} \text{ if } i \neq i$$
$$p_{i,i} = 1 - \sum_{k:k\neq i} q_{i,k} \min\{1, \pi_k/\pi_i\}. \quad (6)$$

This chain is called the *Metropolis chain for* π *with proposal matrix* Q.

After declassification, the method was published in a paper devoted to the statistical physics of fluids. You may recognize Edward Teller among the authors:

N. Metropolis, A. W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller (1953), Equations of state calculations by fast computing machines, Journal of Chemical Physics **21**, 1087-1092.

Markov Chain Monte Carlo

The algorithm of Metropolis

Proposition (Metropolis chain)

Assume that Q is an irreducible symmetric Markov chain on $\mathcal{S}$, and π is a strictly positive probability distribution on $\mathcal{S}$. Then, the Metropolis chain (6) is irreducible and reversible with respect to π and, as a consequence of the detailed balance proposition π is the equilibrium distribution of the Metropolis chain.

Proof

Irreducibility is a direct consequence of irreducibility of Q as $p_{i,j} > 0$ when $q_{i,j} > 0$. If $i = j$ reversibility is trivial. If $i \neq j$, we have the following chain of equalities

$$\pi_i p_{i,j} = q_{i,j}(\pi_i \min\{1, \pi_j/\pi_i\}) = q_{i,j} \min\{\pi_i, \pi_j\} = q_{i,j}(\pi_j \min\{1, \pi_i/\pi_j\}) = \pi_j p_{j,i}. \quad \square$$

Remarks

If the trial state has larger probability than the current state, the trial is always accepted.

For the implementation, we just need to know the ratios π_j/π_i. The normalizing constants for π are not needed. This is very useful for applications.

Markov Chain Monte Carlo

The algorithm of Metropolis

Example: The Ising model

Let $G = (V, E)$ be a graph with vertex set V end edge set E. We use the notation $\langle i, j \rangle = \langle j, i \rangle$ to denote the edge with endpoints at the vertices i and j. To each vertex i, we associate a *spin* variable s_i that can assume either value $s_i = +1$ if the spin is *up* or $s_i = -1$ if the spin is *down*. The state space $\mathcal{S}$ of the Ising model is $\mathcal{S} = \{-1, +1\}^V$. Let $\mathbf{s}$ represent a state in $\mathcal{S}$. The model is characterized by the Hamitonian

$$H(\mathbf{s}) = -J \sum_{\langle i,j \rangle \in E} s_i s_j, \tag{7}$$

where J is a coupling constant ($J > 0$ in the ferromagnetic case, $J < 0$ in the anti-ferromagnetic case and J random for spin glasses) and the probability of a state is given by the distribution of Gibbs

$$\pi(\mathbf{s}; \beta) = \frac{\exp[-\beta H(\mathbf{s})]}{Z(\beta)}, \tag{8}$$

where $Z(\beta)$ is the partition function defined as

$$Z(\beta) = \sum_{\mathbf{s} \in \mathcal{S}} \exp[-\beta H(\mathbf{s})]. \tag{9}$$

For physicists $\beta = 1/kT$, where T is absolute temperature and $k = 1.38 \cdot 10^{-23}\text{JK}^{-1}$ is Boltzmann's constant.

Markov Chain Monte Carlo

The algorithm of Metropolis

Example: The Ising model

From $Z(\beta)$, one can recover the thermodynamics of the system, using the equation

$$F = -kT \log Z, \tag{10}$$

where F is a thermodynamic potential. The entropy S is given by

$$S = -\frac{\partial F}{\partial T}, \tag{11}$$

the heat capacity is

$$C = T\frac{\partial S}{\partial T}. \tag{12}$$

Phase transitions appear as singularities of Z. Given a state $\mathbf{s} \in \mathcal{S}$, one can compute the magnetization $M(\mathbf{s})$ as

$$M(\mathbf{s}) = \sum_{i \in V} s_i \tag{13}$$

The magnetization per spin in state $\mathbf{s}$ is $m(\mathbf{s}) = M(\mathbf{s})/|V|$ and the expected magnetization per spin $m = \mathbb{E}[m(\mathbf{s})]$ is

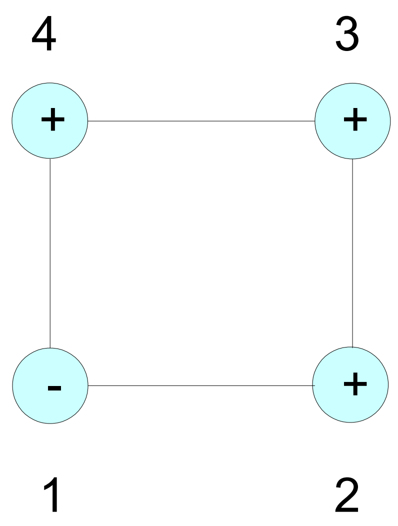
$$m = \sum_{\mathbf{s} \in \mathcal{S}} m(\mathbf{s}) \frac{\exp[-\beta H(\mathbf{s})]}{Z(\beta)}; \tag{14}$$

m is in the interval $[-1, 1]$.

Markov Chain Monte Carlo

The algorithm of Metropolis

Example: The Ising model (square lattice plaquette)



state	$H(\mathbf{s})$	$M(\mathbf{s})$
$++++$	$-4J$	4
$----$	$-4J$	-4
$+++-$	0	2
$++-+$	0	2
$+-++$	0	2
$-+++$	0	2
$++--$	0	0
$+--+$	0	0
$-++-$	0	0
$--++$	0	0
$---+$	0	-2
$--+-$	0	-2
$-+--$	0	-2
$+---$	0	-2
$+-+-$	$4J$	0
$-+-+$	$4J$	0

$$\mathbf{s} = (-1, +1, +1, +1)$$

$$H(\mathbf{s}) = -J(s_1 s_2 + s_2 s_3 + s_3 s_4 + s_4 s_1) = -J(-1 + 1 + 1 - 1) = 0$$

Markov Chain Monte Carlo

The algorithm of Metropolis

Example: The Ising model (square lattice plaquette)

The partition function $Z(\beta)$ is

$$Z(\beta) = 2\,\text{exp}(4\beta J) + 2\,\text{exp}(-4\beta J) + 12 = 4\cosh(4\beta J) + 12 = 4(\cosh(4\beta J) + 3).$$

The thermodynamic potential F is

$$F = -kT\log Z = -\frac{1}{\beta}(\log(\cosh(4\beta J) + 3) + \log(4)).$$

The entropy S is

$$S = -\frac{\partial F}{\partial T} = -k\frac{4\beta J\sinh(4\beta J)}{\cosh(4\beta J) + 3} + k(\log(\cosh(4\beta J) + 3) + \log(4)).$$

The energy E is given by

$$E = -\frac{1}{Z}\frac{\partial Z}{\partial \beta} = F + TS = F + \frac{1}{k\beta}S = -\frac{4J\sinh(4\beta J)}{\cosh(4\beta J) + 3}.$$

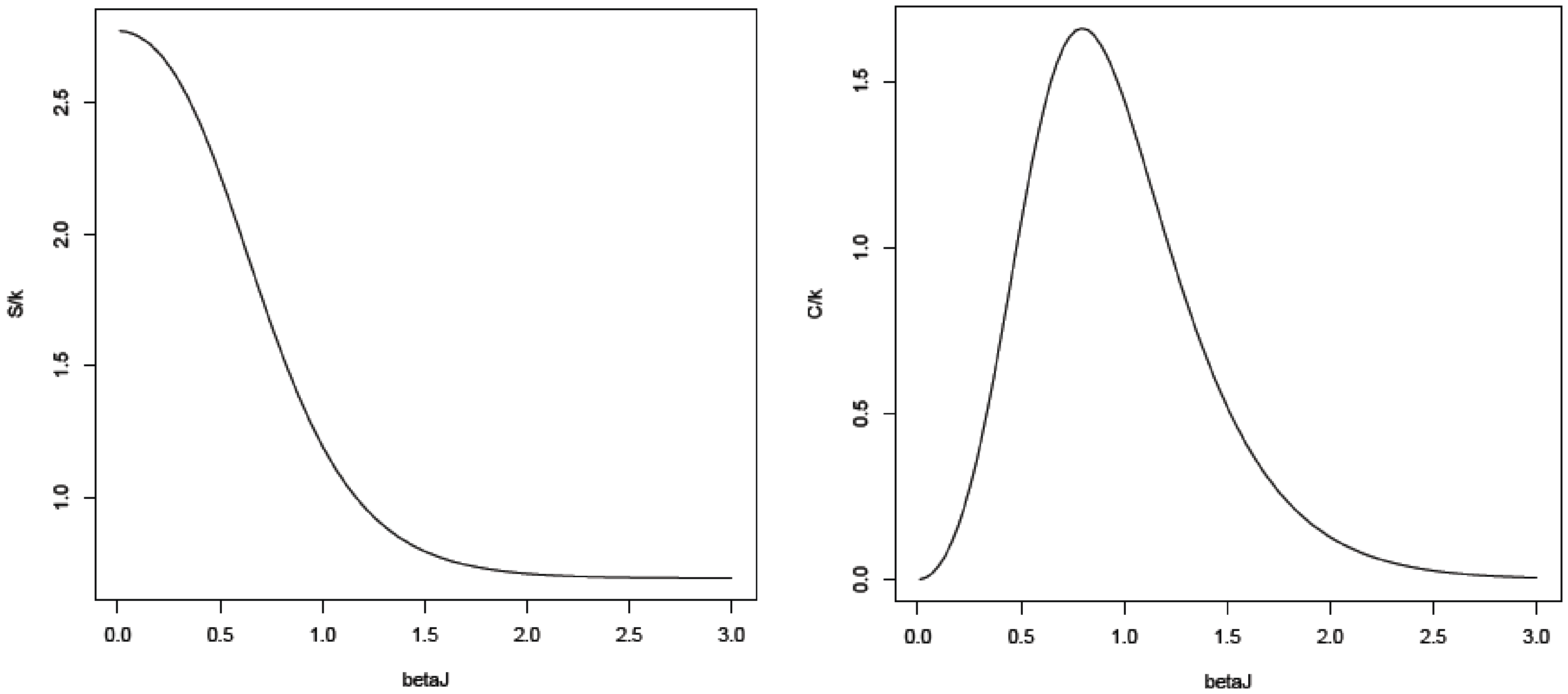
Finally, the specific heat is

$$C = T\frac{\partial S}{\partial T} = \frac{k(4\beta J)^2(3\cosh(4\beta J) + 1)}{(\cosh(4\beta J) + 3)^2}$$

Markov Chain Monte Carlo

The algorithm of Metropolis

Example: The Ising model (square lattice plaquette)



Entropy and specific heat as a function of inverse temperature. The peak in specific heat is a signature of a phase transition in the limit $|V| \to \infty$.

Markov Chain Monte Carlo

The algorithm of Metropolis

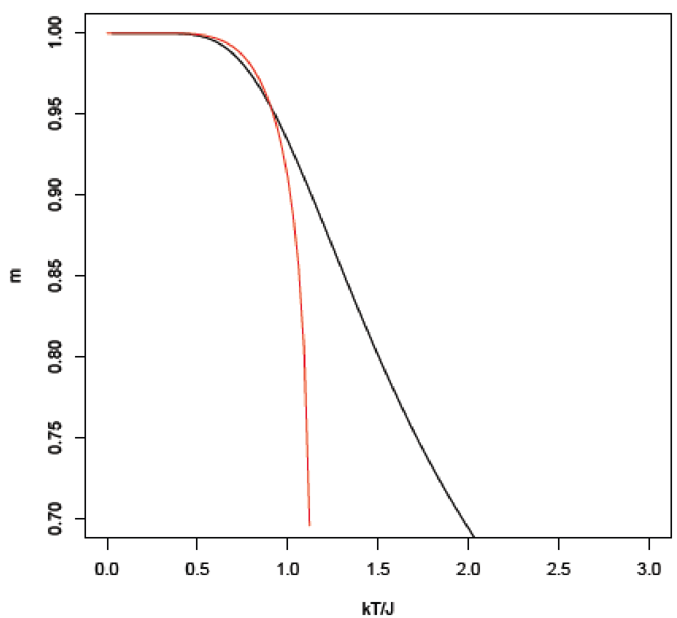
Example: The Ising model (square lattice plaquette)

The average magnetization per spin m vanishes at any finite temperature. The average absolute magnetization $|m|$ is given by

$$|m| = \frac{2\exp(4\beta J) + 4}{4(\cosh(4\beta J) + 3)}.$$

This can be compared with exact Onsager's formula for $kT/J \leq kT_c/J = 1/\log(1+\sqrt{2})$

$$|m| = ((1 - (\sinh(J/kT))^{-4})^{1/8}.$$



Markov Chain Monte Carlo

The algorithm of Metropolis

Example: The Ising model

In order to apply the Metropolis algorithm to the Ising model with target distribution (8), we use the proposal matrix of the simple random walk on $\{-1,+1\}^V$. Here are the steps of the algorithm

1. with the system in state **s**, select a spin $I \in V$ at random and flip it.
2. Set $\mathbf{Y} = (s_1, \ldots s_{I-1}, -s_I, s_{I+1}, \ldots, s_{|V|})$ and $\alpha = \min\{1, \exp[-\beta(H(\mathbf{Y}) - H(\mathbf{s}))]\}$.
3. Generate $U \sim U[0,1]$. If $U \leq \alpha$ accept the trial move and set $X_{t+1} = \mathbf{Y}$, if not, set $X_{t+1} = \mathbf{s}$.

Remarks In step 2, the energy difference only depends on the interactions between s_I and its nearest neighbors. As already mentioned, fortunately, we do not have to compute the partition function (the normalization factor for the Gibbs distribution).

Markov Chain Monte Carlo

The algorithm of Metropolis

Example: The Ising model

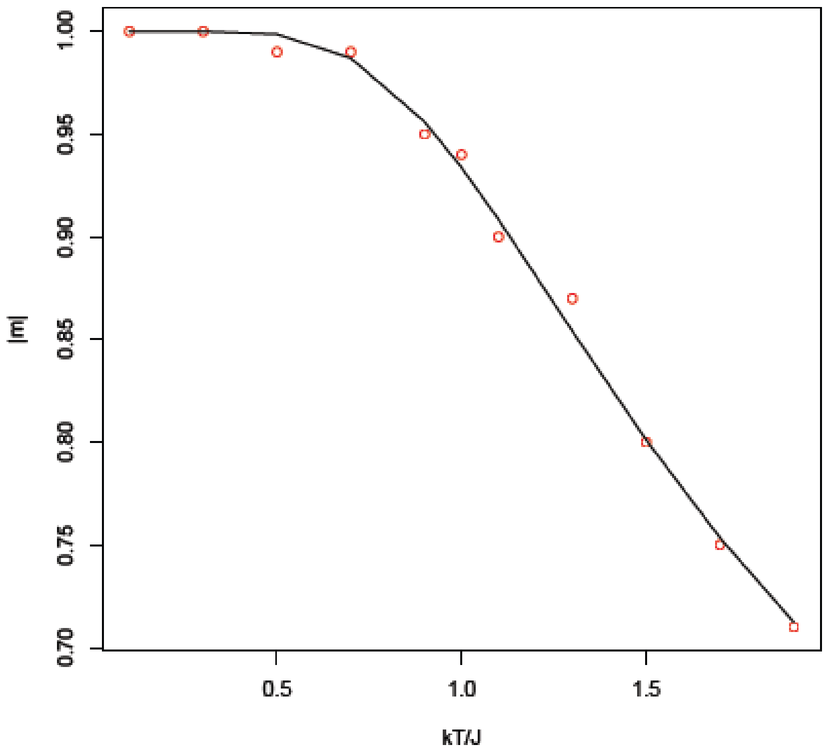
At http://physics.ucsc.edu/~peter/ising/ising.html, it is possible to have a qualitative idea of the behavior of the square-lattice Ising model.

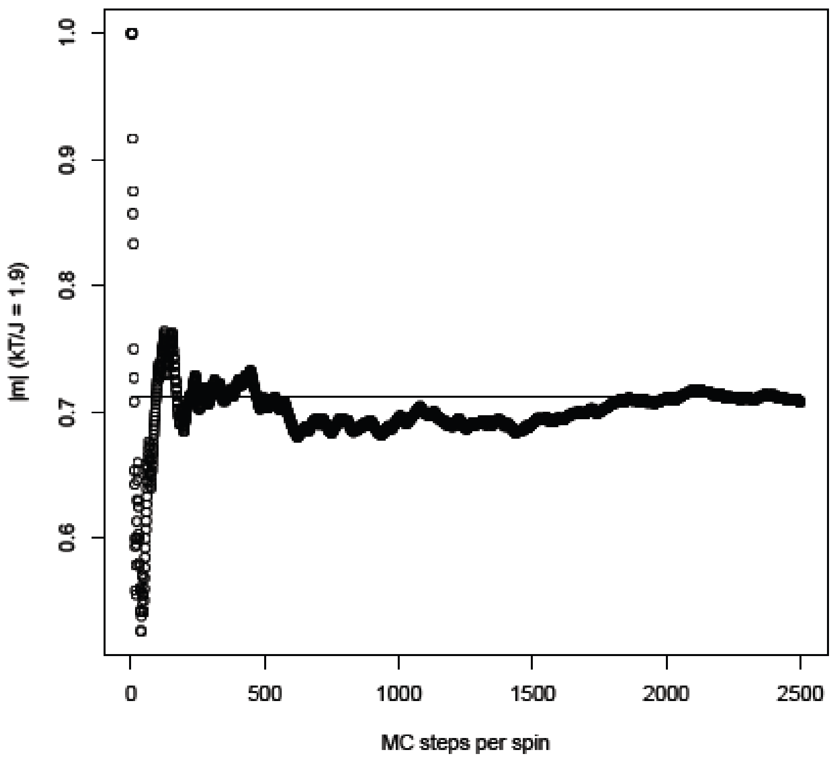
You can see videos on YouTube:
http://www.youtube.com/watch?v=XtY5dLN5CGE (5')
including presentations on how to write a Monte Carlo algorithm:
http://www.youtube.com/watch?v=gsMNqJee87E

Markov Chain Monte Carlo

The algorithm of Metropolis

Example: The Ising model (square lattice plaquette)





Markov Chain Monte Carlo

Final remarks

- Errors on Monte Carlo estimates: Assume you can take a subsample of values that are approximately uncorrelated. Then you can use the usual statistical methods. The estimator of the expected value we have used before is the usual mean, which is unbiased (its expected value coincides with the expected value to be estimated) and consistent (it converges almost surely to the expected value to be estimated). The variance of the mean estimator is proportional to $1/T$, where T is the number of samples. Its standard deviation goes as $1/\sqrt{T}$. This means that to reduce the error of 10 times, you need to run a simulation which is 100 times long. When possible try to use rigorous statistical analysis including correlations.
- Convergence rate of Markov Chain Monte Carlo: There are several results on the speed of convergence of MCMC. Again, when possible, try to use rigorous probabilistic analysis. If this fails, try to see if relevant moments seem to stabilize (has we have done before).

Markov Chain Monte Carlo

Final remarks

- Topics not covered up to now: They are a lot. We have not (yet) covered variance reduction techniques, estimate of integrals with Monte Carlo, other algorithms such as Metropolis-Hastings, the Gibbs sampler or Swendsen-Wang cluster algorithm and many more. Further consider that there many applications of Monte Carlo techniques outside physics. For instance in statistics, Monte Carlo methods are routinely used for goodness of fit tests, for Bayesian estimates, etc.. Elena Akhmatskaya will cover some more advanced topics.

Markov Chain Monte Carlo

It is time for me to go, thank you and goodbye!