Markov Chain Monte Carlo sampling,

diagnostics, and

an approximate exact method.

by Christopher Jennison,

Tine Møller Sørenson,

Richard Sharp

University of Bath

Outline of Talk

1. MCMC for estimating $E\{g(X)\}$ when $X \sim \pi$

2. MCMC diagnostics

3. The PW exact method

4. Using PW — how many chains?

5. Making PW work for general problems

1. Basics of MCMC Estimation

Problem: To estimate $\theta = E\{g(X)\}$ when $X \sim \pi$

Method:

Create a Markov chain with transition matrix P satisfying

$$\pi P = \pi,$$

so the distribution of $X_n \to \pi$ as $n \to \infty$.

The Metropolis-Hastings and Gibbs sampler do this automatically.

From the sample $X_0, X_1, X_2, \ldots,$ form

$$\widehat{\theta} = \frac{1}{N-B} \sum_{i=B+1}^{N} g(X_i).$$

Our estimate of $\theta = E\{g(X)\}$ is

$$\widehat{\theta} = \frac{1}{N-B} \sum_{i=B+1}^{N} g(X_i).$$

For N large,

$$\widehat{\theta} \approx \theta.$$

Moreover

$$E(\widehat{\theta}\,) \approx \theta \quad \text{and} \quad Var(\widehat{\theta}\,) \approx \frac{\sigma^2}{(N-B)/\tau}\,,$$

where $\sigma^2 = Var_\pi\{g(X)\}$ and

τ can be estimated from $X_0, X_1, X_2, \ldots$.

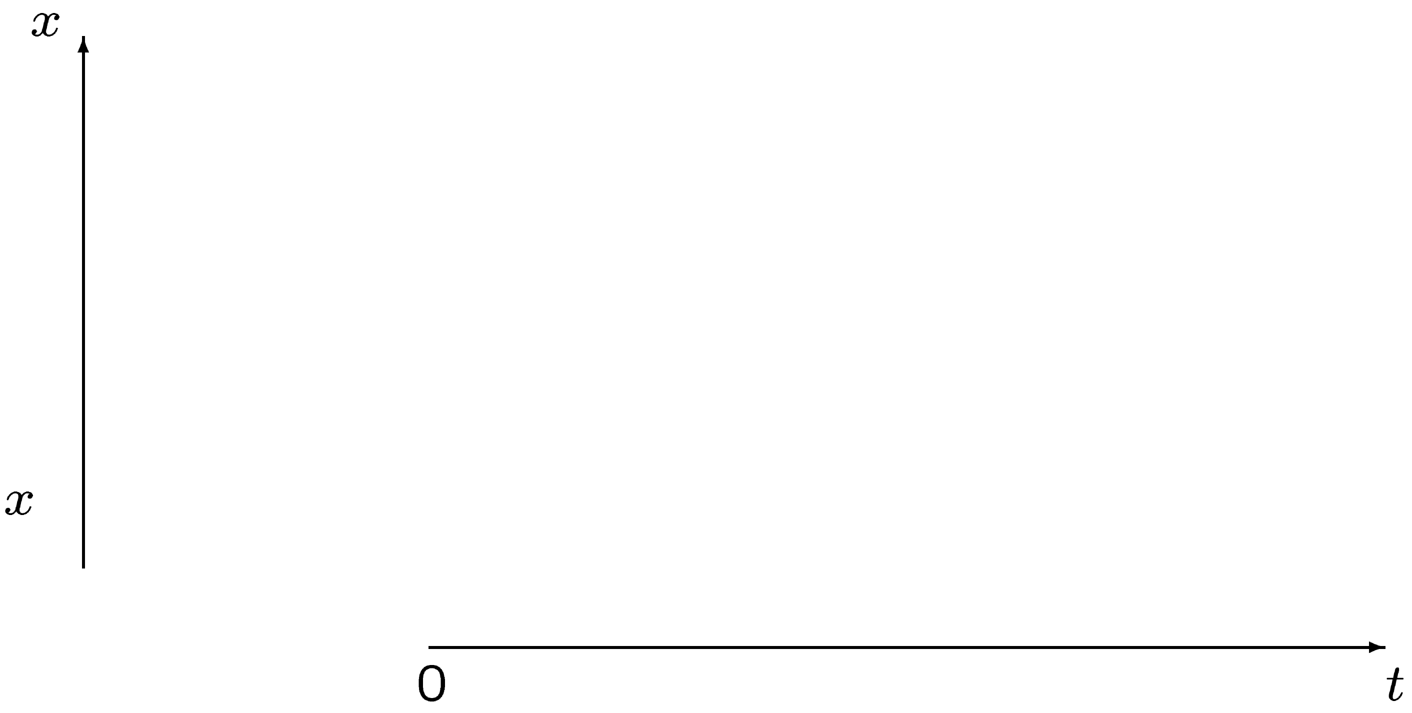
2. Checking that MCMC has Worked

Diagnostics can help assess:

Burn in — have we found the main part of π?

Convergence — has the chain “forgotten” its starting point?

Variance of $\widehat{\theta}$ — how long to run the chain?



Surveys of MCMC diagnostics:

Cowles & Carlin, *JASA*, 1996,

Brooks & Roberts, *Statist. & Comp.*, 1998.

Diagnostics often involve

- multiple chains
- “over-dispersed” starting points
- “coupling” of two or more chains

Even so, you may fail to visit an important part of the sample space — and not know you have missed it.

Assessing convergence is a difficult problem with, as yet, no general solution.

Judging Convergence

Assess the difference between π and the distribution of X_n.

Try to ensure this difference is small at the end of “burn-in”.

Note

1. In calculating

$$\widehat{\theta} = \frac{1}{N-B} \sum_{i=B+1}^{N} g(X_i)\,,$$

early errors are down-weighted by later data.

2. Even if $X_n \sim \pi$ exactly, the process must “forget its history” repeatedly for $\widehat{\theta}$ to be an accurate estimate of θ.

Gelman & Rubin, *Statistical Science*, 1992

Start a number of chains (e.g., 10) from values which are "over-dispersed" relative to π.

Compare within-chain and between-chain variation of a scalar quantity of interest.

Stop when small between-chain variation shows the initial "over-dispersion" has been lost.

Drawbacks

1. Difficulties in finding over-dispersed starts?
2. The approach is univariate.
3. A single long chain would be more efficient.

Geyer, Statist. Sci., 1992:

> "Multiple starts are not necessary in practice
> and not sufficient for good practice."

Valen Johnson, *JASA*, 1996

Start c chains at values sampled from P_0, an over-dispersed distribution relative to π.

Couple the chains by using the same uniform random variables in the Gibbs sampler — this will promote convergence.

Let

$N =$ time at which all c chains converge,

$1 - r = Pr${Draw from P_0 is accepted} using proposals from P_0 in rejection sampling of π,

$\mathcal{L}(x_t) =$ law of X_t (same for all chains).

Then

$$Pr\{N > t\} \quad \geq \quad (1 - r^c)\,\frac{1}{2}\;\| \, \mathcal{L}(x_t) - \pi \, \| \;\; .$$

Johnson, *JASA*, 1996 continued

We can bound $\parallel \mathcal{L}(x_t) - \pi \parallel$ using

$$\parallel \mathcal{L}(x_t) - \pi \parallel \;\leq\; 2 \; Pr\{N > t\} \, \frac{1}{1 - r^c}$$

and an estimate of $Pr\{N > t\}$.

E.g., use multiple runs of coupled chains to find upper percentiles of the distribution of N.

Estimating r needs knowledge of π — from long-run samples, believed close to convergence.

Note:

We assume the method is working properly in implementing the diagnostic procedure — fine, since a rough estimate of π suffices here.

But, don't expect to be warned if the sampler has missed a mode completely.

Johnson, *JASA*, 1996, ...

In examples, Johnson uses results from multiple runs of coupled chains to find a value of N at which one can expect all chains to have coupled.

He then recommends use of this N in setting the duration of burn-in for one long production run.

Efficiency is not really discussed.

3. Propp & Wilson’s Exact Method

Reference: Propp & Wilson, *Random Structures
and Algorithms*, 1996.

First, a demonstration:

Professor Dynkin’s card trick.

Markov Chain Underlying the Card Trick

At stage n, let

$$X_n \quad = \quad \text{Number of cards to go before you reach the next card to “land on”.}$$

Initially, $n = 0$ and X_n is the place of your chosen card in the top row.

If you choose the 3rd card,

$$\begin{aligned}
X_0 &= 3 \\
X_1 &= \\
X_2 &= \\
X_3 &= \\
X_4 &= \\
X_5 &= \\
X_6 &= \\
X_7 &= \\
X_8 &= \\
X_9 &= \\
X_{10} &=
\end{aligned}$$

Underlying Markov Chain ...

State space is $\{1, 2, \ldots, 10\}$.

$$X_{n+1} = \begin{cases} X_n - 1, & \text{for } X_n \geq 2 \\ \text{Value of next card}, & \text{for } X_n = 1. \end{cases}$$

For simplicity, assume an infinite deck of cards.

The transition matrix is $P =$

$$\begin{pmatrix} \frac{2}{13} & \frac{2}{13} & \frac{2}{13} & \frac{1}{13} & \frac{1}{13} & \frac{1}{13} & \frac{1}{13} & \frac{1}{13} & \frac{1}{13} & \frac{1}{13} \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

where

$$\begin{aligned} P_{ij} &= P\{X_{n+1} = j \mid X_n = i\} \\ &= P\{X_{n+1} = j \mid X_n = i, X_{n-1} = x_{n-1}, \ \ldots\}. \end{aligned}$$

Why the Card Trick Works

Let two people, A and B, have Markov chains

$$\{A_n;\, n = 0, 1, \ldots, 52\} \text{ and}$$

$$\{B_n;\, n = 0, 1, \ldots, 52\}.$$

If these meet at n^*, then $A_n = B_n$ for all $n \geq n^*$.

The trick works if chains from all 10 initial states converge before the cards run out.

Dynkin has shown this happens with probability $1 - \epsilon$, where ϵ is very small indeed.

The chains $\{A_n\}$ and $\{B_n\}$ are *coupled* as they use the same random numbers to choose their transitions.

An Application

The limiting distribution of the card trick's Markov chain is the solution, π, to

$$\pi P = \pi. \tag{1}$$

In fact, $\pi = \frac{1}{61}(13, 11, 9, 7, 6, 5, 4, 3, 2, 1)$.

Claim: If we deal 52 cards from an infinite deck, choose an arbitrary X_0 and follow the card trick rules, then $X_{52} \sim \pi$.

Proof:

Imagine we had generated our X_0 from π.

Then, by (1),

$$X_1 \sim \pi P = \pi, \quad \ldots, \quad X_{52} \sim \pi P = \pi.$$

Since all X_0 s lead to the same X_{52}, we can pretend we *did* generate our X_0 from π.

If we can attend to the possibility that chains may not converge in time, we shall have a method for

exact simulation from the past.

Propp & Wilson's algorithm

Define the usual type of Markov chain with limit distribution π.

Run this chain from a time *before* $t = 0$ to produce an $X_0 \sim \pi$ exactly.

Key ingredients

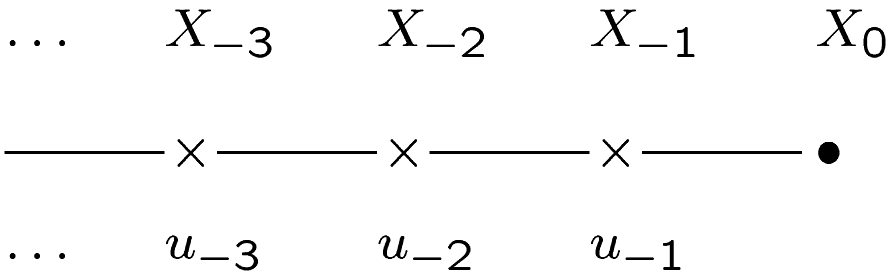
1. Coupling chains in "pre-sampling"

2. The stationarity of π with respect to P:

   if $X_{-n} \sim \pi$ and the transition matrix from $-n$ to $-n+1$ is P, then X_{-n+1} has distribution

$$\pi P \ = \ \pi.$$

Coupling at Times $t < 0$



Here, u_{-1}, u_{-2}, $\ldots$, are realisations of $U(0,1)$ random variables.

Transitions are given by

$$X_{-n+1} \;=\; f(X_{-n}, u_{-n})$$

— once the u_{-n} are specified, f is deterministic.

Chains from different starting points are coupled since they use the same $\{u_{-n}\}$.

When two chains meet up, they stay together.

Using the Stationarity of π

$$X_{-n_1}, \ldots, X_{-1}, \qquad X_0$$

Run coupled Markov chains, all using the same random numbers, from time $-n_1$ to time 0.

Check if chains from all states X_{-n_1} converge.
(A smart way to check this will be useful !)

Suppose all chains do converge:

> We can *claim* to have taken $X_{-n_1} \sim \pi$ so, by stationarity of the Markov chain,
>
> $$X_0 \sim \pi.$$

If all chains do not converge, we must find a way actually to generate $X_{-n_1} \sim \pi$.

The Propp & Wilson Scheme

$$\ldots, \quad X_{-n_2}, \ldots, X_{-n_1-1}, \quad X_{-n_1}, \ldots, X_{-1}, \quad X_0$$

Run coupled Markov chains using the same random numbers from $-n_1$ to 0.

Check if chains from all states X_{-n_1} converge.

```
If so:
    take $X_0$ as a sample from $\pi$
if not:
    sample from time $-n_2$ to time $0$, using
    previous random numbers from $-n_1$ to 0.
    if still no convergence:
        go further back, . . .
```

Back to the Future!

Formal Justification

Consider chains from time $-N$ with $X_{-N} \sim \pi$ and transitions according to P.

We couple chains from different values of X_{-N} so that they stay together once they meet up.

We note transitions in the period $-N$ to 0 for incorporation in chains from earlier starts.

Now suppose

$$P\{\text{Chains from all values } X_{-N} \text{ converge in } N \text{ steps}\} \rightarrow 1 \quad \text{as } N \rightarrow \infty. \tag{2}$$

Let $X_0^{(N)}$ be the final state of one chain from $X_{-N} \sim \pi$.

By (2), $\lim_{N\to\infty} X_0^{(N)}$ exists with probability 1. Call this $X_0^{(\infty)}$.

Claim: $X_0^{(\infty)} \sim \pi$.

Proof: Given $\epsilon > 0$, take N such that

$$P\{X_0^{(\infty)} = X_0^{(N)}\} \geq 1 - \epsilon. \tag{3}$$

The chain yielding $X_0^{(N)}$ has distribution π at time $-N$, and by the equilibrium property of π, we also have $X_0^{(N)} \sim \pi$.

Thus, by (3),

$$|P\{X_0^{(\infty)} = j\} - \pi_j| \;=\; |P\{X_0^{(\infty)} = j\} - P\{X_0^{(N)} = j\}| \;\leq\; \epsilon.$$

Checking Convergence

We wish to check whether chains from all initial states converge in the period $-n_1$ to 0.

PW give a method for problems which have a partial ordering with two extreme states, and MC transitions preserve this ordering.

All chains are sandwiched between the two extremal chains. When these meet, all chains must have converged.

If the two extremal chains do not converge in time $-n_1$ to 0, try from time $-n_2$, etc.

Implementation: Note random number seeds at times $-n_1$, $-n_2$, $\ldots$. If chains revisit one of these times reset the seed to its previous value.

Example: Ising Model on an $\mathbf{n} \times \mathbf{n}$ Lattice

Each element $X(i,j)$ of the random variable

$$X = \{X(i,j);\ i = 1,\ldots,n,\ j = 1,\ldots,n\}$$

takes a value 0 or 1. The sample space is

$$\Omega = \{0,1\}^{n^2}.$$

In the 4-neighbour model with parameter β

$$Pr\{X{=}x\} = \frac{1}{Z}\exp[-\beta \sum_{\langle ij,kl\rangle} I\{x(i,j) \neq x(k,l)\}],$$

where $\langle ij,\, kl\rangle$ indicates summation over indices i,j and k,l for which $X(i,j)$ and $X(k,l)$ are horizontal or vertical neighbours.

Conditionally, given all $X(k,l)$, $(k,l) \neq (i,j)$,

$$X(i,j) = \begin{cases} 0 & \text{with prob. } e^{\beta n_0}/(e^{\beta n_0} + e^{\beta n_1}) \\ 1 & \text{with prob. } e^{\beta n_1}/(e^{\beta n_0} + e^{\beta n_1}), \end{cases}$$

where n_0 and n_1 are numbers of neighbours of $X(i,j)$ equal to 0 and 1, respectively.

A Coupled Gibbs Sampler for the Ising Model

To update $X(i,j)$:

Draw a $U(0,1)$ random variate — call this u.

We want $X(i,j) = 0$ or 1 with probabilities

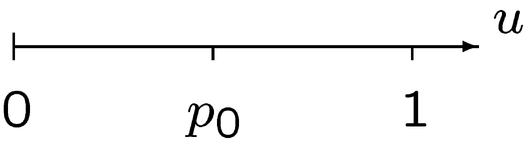
$$p_0 = e^{\beta n_0}/(e^{\beta n_0} + e^{\beta n_1}),$$
$$p_1 = e^{\beta n_1}/(e^{\beta n_0} + e^{\beta n_1}),$$

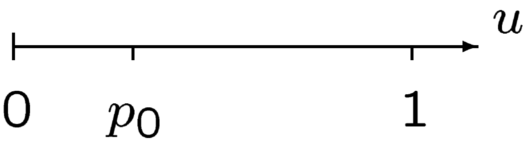
so, set

$$X(i,j) = \begin{cases} 0 & \text{if } u \le p_0, \\ 1 & \text{if } u > p_0. \end{cases}$$

Coupling:

$n_0 = 2,\ n_1 = 2$ $\qquad$ $n_0 = 1,\ n_1 = 3$





Partial ordering:

We preserve a partial ordering with extremes

$\underline{0}$ and $\underline{1}$, the all-0 and all-1 images.

Pre-sampling for the Ising Model

Number of image sweeps needed to converge by time 0, starting at times -5, -10, -20, etc.

Values are means over 40 replicates.

	Lattice size		
β	25×25	50×50	100×100
0.1	5	5	6
0.2	5	9	10
0.3	10	10	12
0.4	13	19	20
0.5	21	28	36
0.6	38	46	60
0.7	71	96	132
0.8	210	332	592
0.9	1148	6770	—
1.0	—	—	—

Blanks indicate failure to converge, usually, in 20480 iterations.

4. How Should We Apply PW?

For the same effort, we could have

1 long chain,

several middling chains, or

many short chains

each with pre-sampling according to PW.

With a fixed total number of iterations to divide between a number of chains:

Efficiency is best with 1 or 2 chains, but not seriously reduced until pre-sampling uses a significant fraction of iterations.

Running several PW chains can *enhance confidence* in the results — a single chain from $X_0 \sim \pi$ could become stuck in part of the sample space

Ising Model, 25 × 25 lattice

With $\beta = 0.5$, mean pre-sample is 73 sweeps.

Simulations use N chains of length L, where

$$N \times (73 + L) = 2000.$$

N	L	$Var(\widehat{\theta}_1)$	$Var(\widehat{\theta}_2)$	$Var(\widehat{\theta}_3)$	$Var(\widehat{\theta}_4)$
		$\times 10^4$	$\times 10^5$	$\times 10^6$	$\times 10^7$
1	1927	2.3	11.7	4.3	2.4
2	927	2.5	12.8	4.4	2.5
3	593	2.6	12.8	4.6	2.7
4	427	2.7	13.2	4.8	2.7
5	327	2.9	14.0	5.1	2.8
10	127	3.7	17.8	6.6	3.6
15	60	5.2	25.1	9.3	5.4
20	27	8.5	40.9	14.2	8.8
25	7	23.0	116.0	30.8	24.1

Efficiency: No serious loss until L approaches
the mean pre-sample length, 73.

Advantages of several chains:

Confidence in $\widehat{\theta}$ — a single chain from $X_0 \sim \pi$ may remain stuck in part of the sample space.

Direct estimation of variance of $\widehat{\theta}$ (?)

5. General Use of the PW Algorithm

We need to establish that chains from all states in Ω converge from time $-T$ to 0.

P & W used chains from 2 extremal states to "sandwich" all other chains.

Some ingenious extensions:

Kendall, *Probability Towards 2000*, 1998,

Häggström, van Lieshout & Møller, *Bernoulli*, to appear,

Murdoch & Green, *Scand. J. Statist.*, 1998.

or visit

`http://dimacs.rutgers.edu/~dbwilson/exact.html/`

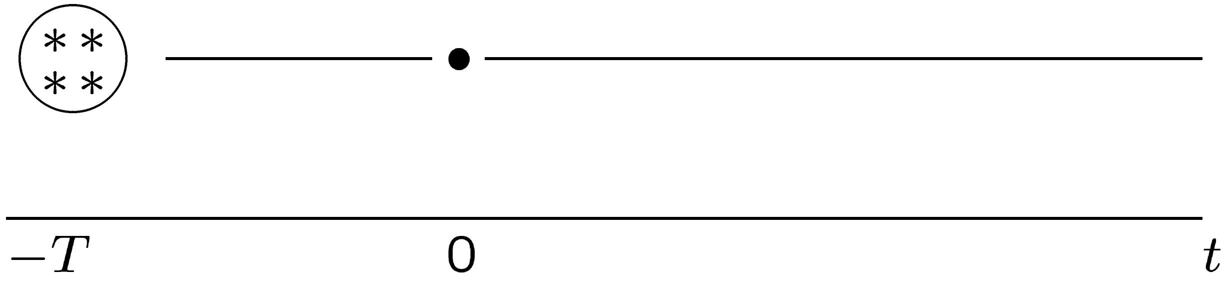
What if this approach fails?

e.g., Potts model (c-colour Ising model),

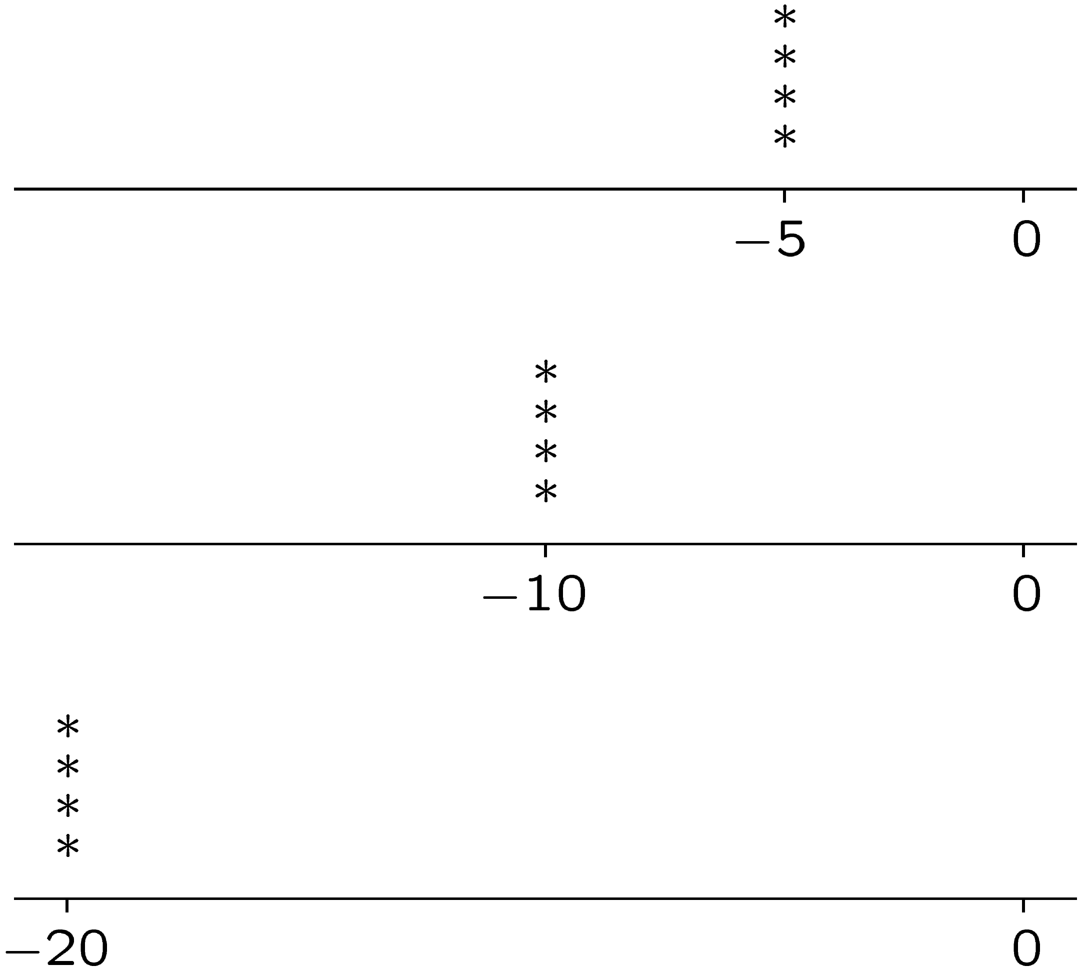
most (?) common problems.

Plan A

Check convergence of K test images.



Pre-sampling:



An Application of Plan A

Example: Ising model, 25 × 25 lattice, $\beta = 0.5$

1. Generate a “test set” of K random images.
   Work back to a $-T$ for which test chains
   converge during the period $-T$ to 0.

2. Check if T is *really* large enough by running
   chains from $\underline{0}$ and $\underline{1}$ over $-T$ to 0.

Results from 100 replicates

	K	% Success
$\gamma = 0.5$	2	51
	5	77
	10	87
$\gamma = 0.7$	2	79
	5	92
	10	99

In alternate test images,

$Pr\{X(i,j) = 0\} = \gamma \; \forall \, i, \, j,$

$Pr\{X(i,j) = 1\} = \gamma \; \forall \, i, \, j.$

What Should We Put in a Test Set?

It seems wise to choose

(1) images from each mode of π, and

(2) images sampled from π.

1) Hope that chains starting within a mode meet

up while waiting for a jump between modes.

2) In justifying the PW method, we argued

“It is *as if* we sampled X_{-T} from π.
Since $\pi P = \pi$, each subsequent $X_t \sim \pi$.
Hence, $X_0 \sim \pi$.”

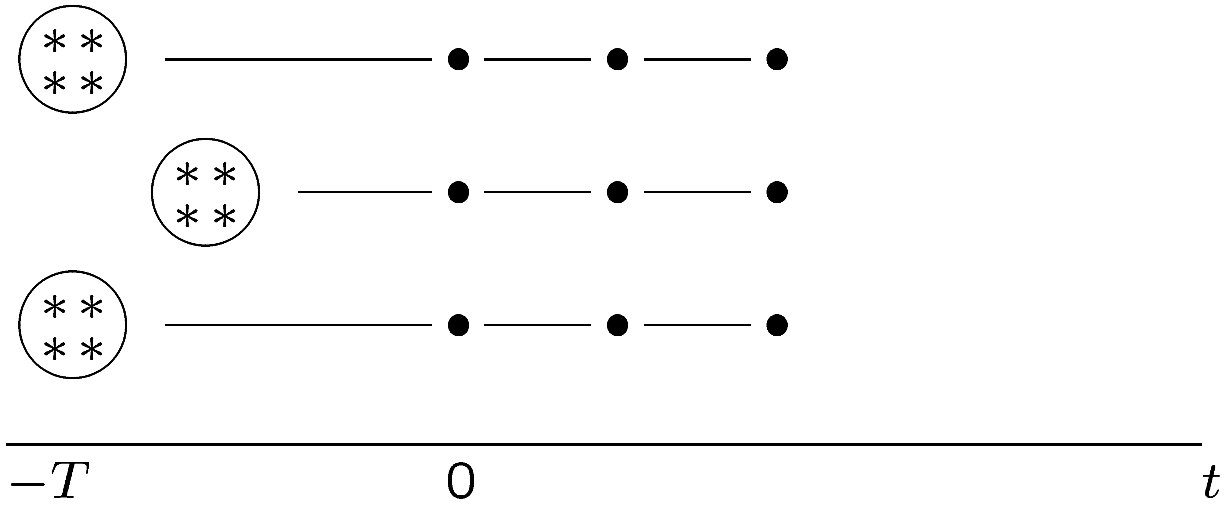
So, we need to be confident that values X_{-T} sampled under π are likely to lead to our X_0.

We can sample MCs beyond $t = 0$ to obtain an approximate sample from π.

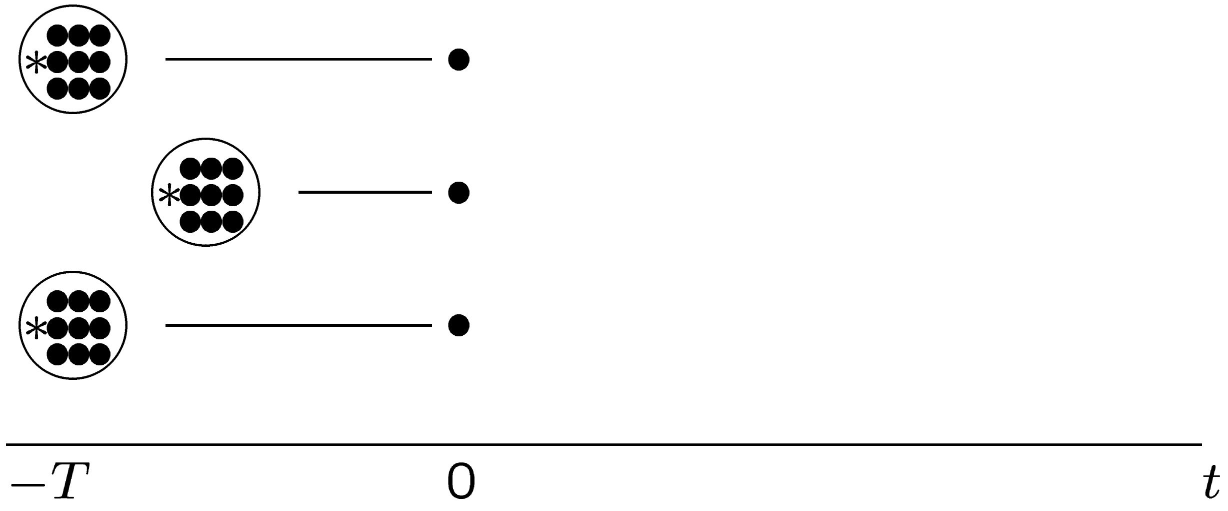
This gives an iterative method for making a suitable “initial test set” — and suggests a diagnostic for later use.

Plan B

Stage 1 — using random test sets



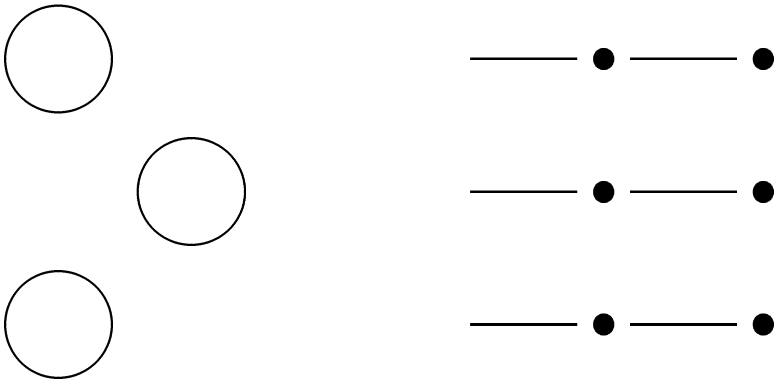
Stage 2 — test sets $\sim \pi$ approx.

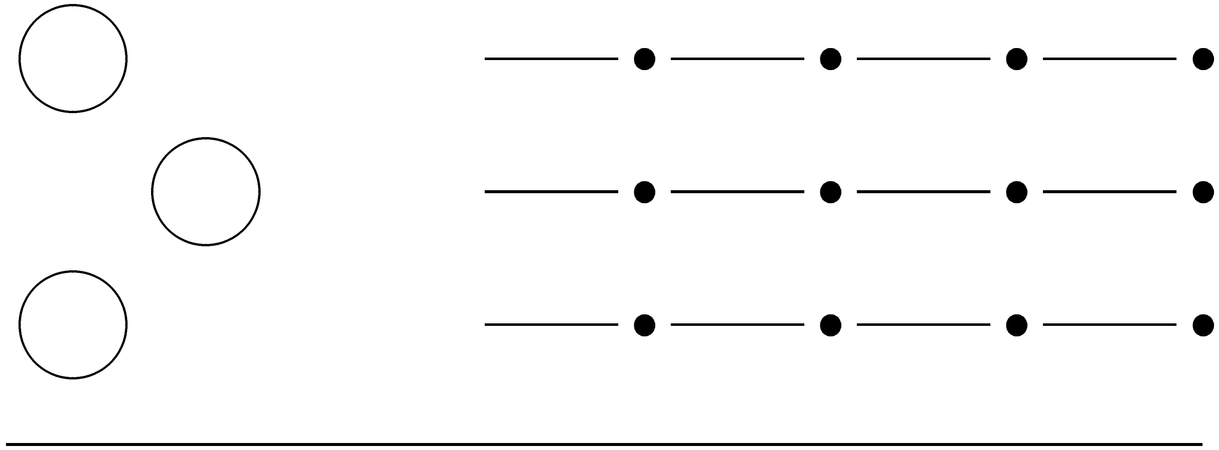


Has each chain converged to its previous X_0 ?

If so — move on to the “production run”.

If not — extend pre-sampling before current $-T$,
re-generate samples. Iterate over Stage 2.





Ising model, 25×25 lattice, $\beta = 0.5$

Results from 50 replicates

		DIAGNOSTICS		
No.	Images		When not OK	
chains	in first	All OK ?	No. chains	Chains →
N	test set		not conv.	a new X_0
2	10	47 (44)*	1.0 (1.0)	0 (0)
2	50	49 (48)	1.0 (1.0)	0 (0)
5	10	48 (46)	1.0 (1.25)	0 (0)
5	50	50 (49)	– (1.0)	0 (0)
10	10	50 (48)	– (1.0)	0 (0)
10	50	49 (46)	1.0 (1.0)	0 (0)

( )* testing convergence from $\underline{0}$ and $\underline{1}$
— a definitive test in this example.

Usually 1 iteration of Stage 2, occasionally 2.

2nd test set: 5 samples from each of N chains
at intervals of 20 iterations.

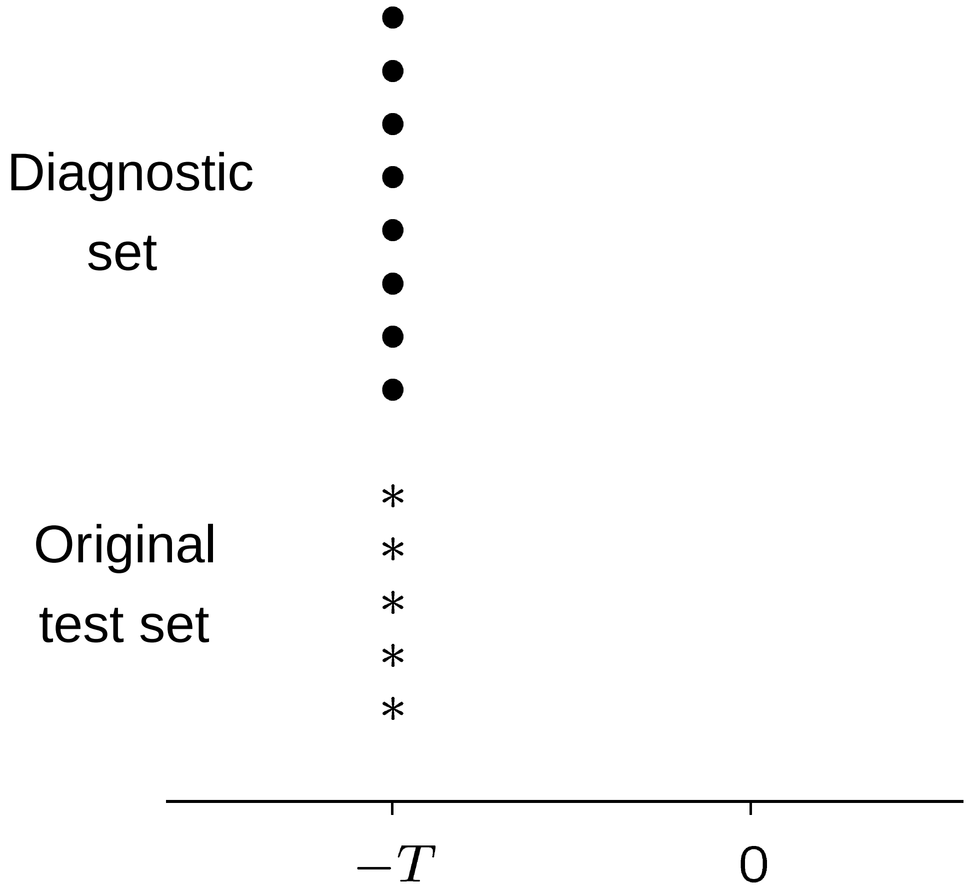
Diagnostics: Test set of 5 samples from the
N chains' production runs at
intervals of 400 iterations.

Response to the final diagnostic:

We could use a longer pre-sampling period
— but results would not change.

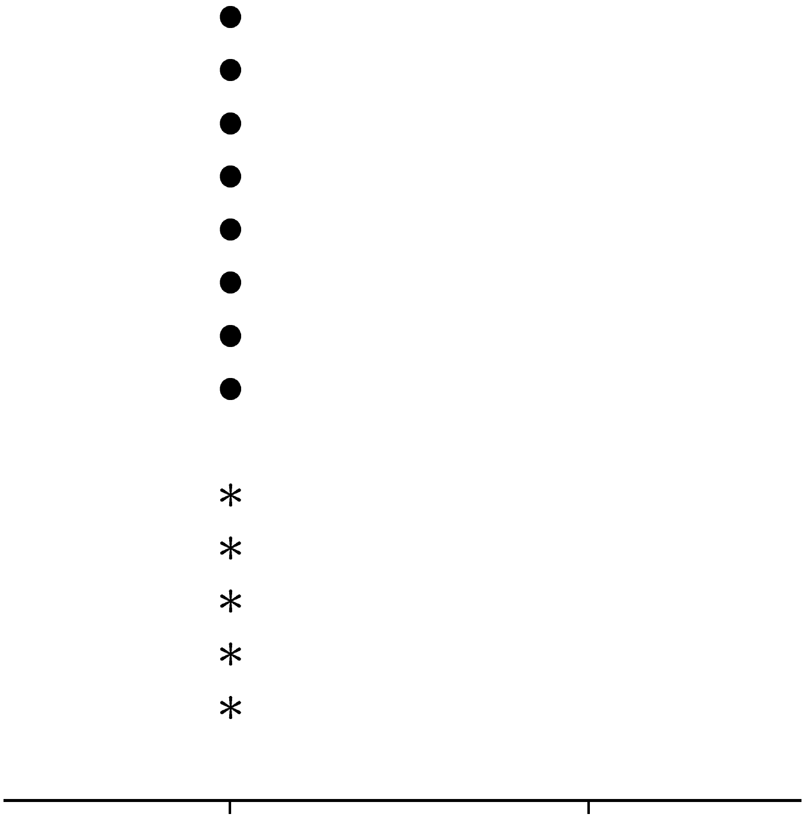
The Final Diagnostic

Looking at one of the N chains:



An error is detected when a new test chain does not meet up by time 0 with all original test chains.

Ideally, we should resolve this by pre-sampling from an earlier point — this may well yield the same X_0.



A Formal Error Bound

Suppose we have chosen $-T$ and our test chains lead from their values of X_{-T} to $X_0 = x_0$.

For a proper sample from π at time 0, we should create $X_{-T} \sim \pi$ and follow *this* chain to time 0.

Define

$$\epsilon = Pr\{X_{-T} \sim \pi \text{ does not lead to } X_0 = x_0\}.$$

With M diagnostic runs from $X_{-T} \sim \pi$, approx,

$$\begin{aligned} & Pr\{\text{All } M \text{ diagnostic runs lead to } x_0 \text{ but} \\ & \qquad \text{the one sampling run does not}\} \\ & \quad = \ (1-\epsilon)^M \, \epsilon \\ & \quad \leq \ (1 - \frac{1}{M+1})^{M+1} \, \frac{1}{M} \\ & \quad < \ \frac{1}{M\, e} \, . \end{aligned}$$

An Error Bound . . .

Allow probability δ that our X_0 is not what we would have obtained doing things "properly".

Overall strategy:

1. Choose a value for $-T$ such that we believe all chains converge between $-T$ and 0.

   Run M_1 diagnostic runs, where

   $$\frac{1}{M_1\, e} = \frac{\delta}{2}\,.$$

   If all M_1 chains converge, accept their common x_0 as a sample from π. If not, . . .

2. Choose an earlier value for $-T$.
   Run M_2 diagnostic runs, where

   $$\frac{1}{M_2\, e} = \frac{\delta}{4}\,,$$

   etc.

5-colour Potts model, 25×25, $\beta = 0.8$

Results from 50 replicates

		DIAGNOSTICS		
No.	Images		When not OK	
chains	in first	All OK ?	No. chains	Chains $\rightarrow$
N	test set		not conv.	a new X_0
2	10	49 (48)*	1.0 (1.0)	0 (0)
2	50	50 (48)	– (1.0)	0 (0)
5	10	49 (46)	1.0 (1.0)	0 (0)
5	50	48 (43)	1.0 (1.0)	0.5 (0.14)
10	10	47 (44)	1.0 (1.0)	0 (0)
10	50	47 (41)	1.0 (1.22)	0 (0)

( )* testing convergence from
5 single-colour images
— not a definitive test.

Usually 1 iteration of Stage 2, occasionally 2 or 3.

2nd test set: 5 samples from each of N chains
at intervals of 20 iterations.

Diagnostics: Test set of 5 samples from the
N chains' production runs at
intervals of 400 iterations.

Response to the final diagnostic:

We could use a longer pre-sampling period
— affecting one chain in one replication.

10-colour Potts model, 25×25, $\beta = 0.5$

Results from 50 replicates

		DIAGNOSTICS		
No.	Images		When not OK	
chains	in first	All OK ?	No. chains	Chains →
N	test set		not conv.	a new X_0
2	10	46 (44)*	1.0 (1.0)	0 (0.17)
2	50	50 (50)	– (–)	0 (0)
5	10	47 (37)	1.0 (1.0)	0 (0)
5	50	46 (45)	1.0 (1.0)	0 (0)
10	10	47 (34)	1.0 (1.25)	0 (0.06)
10	50	48 (36)	1.0 (1.0)	0 (0)

( )* testing convergence from
5 single-colour images
— not a definitive test.

Usually 1 or 2 iterations of Stage 2, occasionally 3.

2nd test set: 5 samples from each of N chains
at intervals of 20 iterations.

Diagnostics: Test set of 5 samples from the
N chains' production runs at
intervals of 400 iterations.

Response to the final diagnostic:

We could use a longer pre-sampling period
— affecting one chain in two replications.

10-colour Potts model, 25×25, $\beta = 0.5$

Results from 50 replicates of the modified method, using earliest $-T$ for *all* chains.

		DIAGNOSTICS		
			When not OK	
No. chains N	Images in first test set	All OK ?	No. chains not conv.	Chains → a new X_0
2	10	49 (49)*	1.0 (2.0)	0 (0)
2	50	50 (50)	– (–)	0 (0)
5	10	50 (49)	– (1.0)	0 (0)
5	50	50 (50)	– (–)	0 (0)
10	10	50 (50)	– (–)	0 (0)
10	50	50 (50)	– (–)	0 (0)

( )* testing convergence from
5 single-colour images
— not a definitive test.

Conclusions and Discussion

Our algorithm is effective in examples:

It provides approximately exact samples and backs these up with diagnostics.

Computational needs:

Pre-sampling and checking can take roughly similar effort to “production” samples — the price of confidence in MCMC results.

When might the algorithm fail?

If one mode is never seen — a pitfall for all diagnostics.

What do you need to use it?

Usual Markov chains — plus coupling.