Convergence rates for MCMC

About convergence for Markov chains

When using Markov chains as a simulation tool it is important that s stationary distribution exists- Markov chain Monte Carlo and the Gibbs sampler, or the Metropolis chain, helps us generate samples from any distribution stationary for the chain.

There are two problems we have to adress . . .

(A) The distribution of X_n might never become equal to π, regardless how large we let n be. We will however get close in total variation meaning.

(B) We do not know is how fast $\mu^{(n)}$ approaches the stationary distribution π. We would like some result telling us

$$d_{\mathrm{TV}}(\mu^{(n)}, \pi) \leq \varepsilon \quad \Rightarrow \quad n \geq g(\varepsilon)$$

For some function g of ε. In general there exists no bounds of practical importance even though some theoretical bounds exists.

Of these two B is the most serious one. For Markov chains having finite state spaces we can say something about the total variation distance between $\mu^{(n)}$ and π.

1. $d_{\mathrm{TV}}(\mu^{(n)}, \pi)$ is a decreasing function of n.
2. d_{TV} decays exponentially, that is,

$$d_{\mathrm{TV}}(\mu^{(n)}, \pi) \leq C_1 e^{-C_2 n}$$

for some $C_1 < \infty$ and some $C_2 > 0$.

The first one is not a surprise. The second statement could be useful if C_1 isn't too large and if C_2 isn't too close to 0, but of course, in general they are.

Something about convergence modes

There are some examples of how the total variation distance decreases as n grows larger. Today there are no characterisation results, meaning that there are no results telling

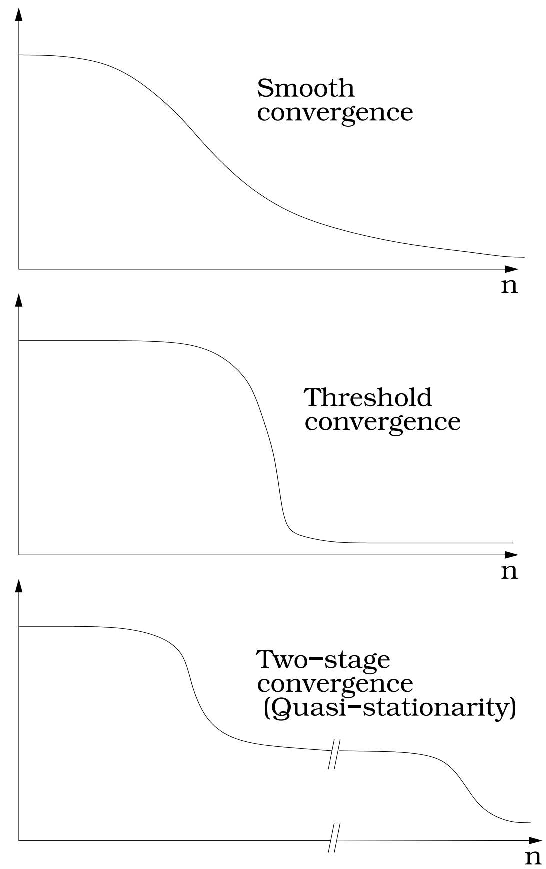

Figure 8.1: Three examples of how the total variation distance decreases as n increases.

us what mode of convergence a certain Markov chain have. In an application any of these three can apply. For some special cases there has been progress, and the mode of convergence has been determined, but in general we do not know.

Fast convergence of the MCMC algorithm for generating random q-colorings

The systematic sweep Gibbs sampler

A version of the Gibbs sampler for random q-colorings is the following. We order the vertices in V is some order.

$$V = \{v_1, v_2, ..., v_k\}$$

Instead of choosing a vertex at random to update we update the in order $v_1, v_2, ..., v_k$. This makes the Markov chain time inhomogenous, but it still has $\rho_{G,q}$ as reversible distribution. Further this sampler is irreducible if the "random vertex" Gibbs sampler is irrducible, and we have shown that it is. So indeed we still have convergens of $\mu^{(n)}$ towards $\rho_{G,q}$.

Bounding total variation distance using coupling

Coupling can be used to bound the distance of total variation. It is used in the proof of theorem 8.1 but in a general technique. The idea is the following. We contruct a coupling of two random variables Y_1 and Y_2 and bound the total variation distance by the probability that they are unequal. First we establish the inequality (without using any coupling).

Let pi_1 and π_2 be two probability distribution on some finite set S. Let Y_1 and Y_2 be two random variables with distributions π_1 and π_2 respectively. Consider the definition of the total variation distance.

$$d_{\text{TV}}(\pi_1, \pi_2) = \max_{\{A : A \subseteq S\}} |\pi_1(A) - \pi_2(A)| = \max_{\{A : A \subseteq S\}} |\mathbb{P}(Y_1 \in A) - \mathbb{P}(Y_2 \in A)|$$

For any such event A we have the following.

$$\begin{aligned}
\mathbb{P}(Y_1 \in A) - \mathbb{P}(Y_2 \in A) &= [\ \mathbb{P}(Y_1 \in A, Y_2 \in A) + \mathbb{P}(Y_1 \in A, Y_2 \notin A)\ ] \\
&\quad -[\ \mathbb{P}(Y_2 \in A, Y_1 \in A) + \mathbb{P}(Y_2 \in A, Y_1 \notin A)\ ] \\
&= \mathbb{P}(Y_1 \in A, Y_2 \notin A) - \mathbb{P}(Y_2 \in A, Y_1 \notin A) \\
&\leq \mathbb{P}(Y_1 \in A, Y_2 \notin A) \\
&\leq \mathbb{P}(Y_1 \neq Y_2)
\end{aligned}$$

We can interchange Y_1 and Y_2 and get $\mathbb{P}(Y_2 \in A) - \mathbb{P}(Y_2 \in A) \leq \mathbb{P}(Y_1 \neq Y_2)$. Putting it all together gives us . . .

$$d_{\text{TV}}(\pi_1, \pi_2) = \max_{\{A : A \subseteq S\}} |\mathbb{P}(Y_1 \in A) - \mathbb{P}(Y_2 \in A)| < \mathbb{P}(Y_1 \neq Y_2)$$

. . . and we are done.

Given the inequality we use the coupling to bound $\mathbb{P}(Y_1 \neq Y_2)$. For a Markov chain we typically construct the coupling such that $\mathbb{P}(X_n \neq X'_n)$ approaches zero as $n \to \infty$.

The statement

Theorem 8.1 ***Convergence rate for the systematic sweep Gibbs sampler for random q-coloring***

Let $G = (V, E)$ be a graph. Let k be the number of vertices in G, and suppose that any vertex $v \in V$ has at most d neighbours. Suppose furthermore that $q > 2d^2$. Then, for any fixed $\varepsilon > 0$, the number of iterations needed for the systematic sweep Gibbs sampler to come within total variation distance ε of the target distrubition $\rho_{G,q}$ is at most

$$k \left(\frac{\log(k) + \log(\varepsilon^{-1}) - \log(d)}{\log\left(\frac{q}{2d^2}\right)} + 1 \right)$$

Proving theorem 8.1

Let $G = (V, E)$ be a graph and let $q > 2d^2$. Also let $\rho_{G,q}$ be the measure for random q-colorings of G.

We need to establish a lower bound on n ensuring us that the total variation distance between the distribution on X_n and $\rho_{G,q}$ is small. We start with determine a function $f(n)$ nsuch that

$$d_{\mathrm{TV}}(\mu^{(n)}, \pi) \le f(n)$$

and from this deduce how large n needs to be.

The main tool is a coupling of two chains $(X_0, X_1, ...)$ and $(X'_0, X'_1, ...)$ such that we can bound $\mathbb{P}(X_n \neq X'_n)$ from above for any n. The proof concists of three parts.

1. We construct the coupling.
2. Determine upper bound for $\mathbb{P}(X_n(v) \neq X'_n(v))$ for some $v \in V$.
3. Determine upper bound for $\mathbb{P}(X_n \neq X'_n)$.
4. Determine lower bound on n to ensure $d_{\mathrm{TV}}(\mu^{(n)}, \pi) < \varepsilon$

Step 1

We create two Markov chains $(X_0, X_1, ...)$ and $(X'_0, X'_1, ...)$. Let $X_0 = \xi$ for some arbitrary but fixed element $\xi \in \{1, ..., q\}^V$, let X'_0 be distributed according to $\rho_{G,q}$.

During an update at vertex v do the following. Pick a permutation

$$R = (R^1, R^2, ..R^q)$$

of the colors $1, 2, ..., q$ uniformly at random from the set of $q!$ permutations and let v get the first color in R not used by any neighbours of v. Let

$$X_n(v) = R^i \qquad X'_n(v) = R^{i'}$$

where

$$i = \min\{j : X_n(w) \neq R^j_n \text{ for all neighbours } w \text{ of } v\ \}$$

$$i' = \min\{j' : X_n(w) \neq R^{j'}_n \text{ for all neighbours } w \text{ of } v\ \}$$

For each update we need a new permutation, and the coupling is constructed by using the same permutation for both chains. Let $R_0, R_1, ...$ be the sequence of permutations we use. This defines the coupling.

Step 2

Let $v \in V$ be arbitrary but fixed. When updating the color at vertex v in $X_{n+1}(v)$ and $X'_{n+1}(v)$ we call the update a success if $X_{n+1}(v) = X'_{n+1}(v)$, otherwise it's called a failure.

To bound the probability of a failure we partition the color set $\{1, ..., q\}$. The partition consists of three subsets, defined as follows.

B_2 : Colors used among neighbours of v in both $X_n(v)$ and $X'_n(v)$.

B_1 : Colors used among neighbours of v in exactly one of $X_n(v)$ or $X'_n(v)$.

B_0 : Colors not used among neighbours of v in both $X_n(v)$ and $X'_n(v)$.

When choosing a color in the permutation any color in B_2 is discarded. If a color in B_1 turns up before any color in B_0 the update is a failure. If we condition on everything else but v then

$$\mathbb{P}(\text{ failed update }) = \frac{B_1}{B_0 + B_1}$$

We can bound this from above.

$$\frac{B_1}{B_0 + B_1} = \frac{B_1}{q - B_2} \leq \frac{2d - 2B_2}{q - B_2} \leq \frac{2d - B_2}{q - B_2} \leq \frac{2d\left(1 - \frac{B_2}{2d}\right)}{q\left(1 - \frac{B_2}{q}\right)} \leq \frac{2d}{q}$$

After making one sweep in the systematic sweep Gibbs sampler we have the following for any $v \in V$.

$$\mathbb{P}(X_k(v) \neq X'_k(v)) \leq \frac{2d}{q}$$

During the second sweep we will have a successfull update at v if the neighbourhood of v is the same in both chains. A failure can occur again if there is a discrepancy between neighbourhoods in the two chains, and if the update is a failure.

$$\mathbb{P}(\text{ discrepancy }) \leq d\,\frac{2d}{q} = \frac{2d^2}{q}$$

So the probability of a failure during the second update is bounded by the following.

$$\mathbb{P}(\text{ failed update }) = \underbrace{\mathbb{P}(\text{ failed update} \mid \text{discrepancy })}_{\leq \frac{2d}{q}}\ \underbrace{\mathbb{P}(\text{ discrepancy })}_{\leq \frac{2d^2}{q}} \leq \frac{2d}{q}\left(\frac{2d^2}{q}\right)$$

We can iterate this. After m sweeps the probability of having different colors at v is bounded accordingly.

$$\mathbb{P}(X_{mk}(v) \neq X'_{mk}(v)) \leq \frac{2d}{q}\left(\frac{2d^2}{q}\right)^{m-1}$$

Step 3

We can now bound the probability that $X_{mk} \neq X'_{mk}$ by noting that there is enough to have $X_{mk}(v) \neq X'_{mk}(v)$ for some vertex $v \in V$. Since we have (by assumption) k vertices we get the following bound.

$$\mathbb{P}(X_{mk} \neq X'_{mk}) \leq \sum_{v \in V} \mathbb{P}(X_{mk}(v) \neq X'_{mk}(v)) \leq k\,\frac{2d}{q}\left(\frac{2d^2}{q}\right)^{m-1} = \frac{k}{d}\left(\frac{2d^2}{q}\right)^{m}$$

Step 4

We bound the total variation distance by using the coupling.

$$d_{\mathrm{TV}}(\mu^{(mk)}, \rho_{G,q}) \leq \mathbb{P}(X_{mk} \neq X'_{mk}) \leq \frac{k}{d}\left(\frac{2d^2}{q}\right)^{m}$$

If we set the distance to ε we get

$$\leq \frac{k}{d}\left(\frac{2d^2}{q}\right)^{m} = \varepsilon \quad \Leftrightarrow \quad m = \frac{\log(k) + \log(\varepsilon^{-1}) - \log(d)}{\log\left(\frac{q}{2d^2}\right)}$$

In order to hace the total variation distance between $\mu^{(n)}$ and $\rho_{G,q}$ we have to make at least

$$m \geq \frac{\log(k) + \log(\varepsilon^{-1}) - \log(d)}{\log\left(\frac{q}{2d^2}\right)}$$

sweeps, or, at least,

$$n \geq k \left(\frac{\log(k) + \log(\varepsilon^{-1}) - \log(d)}{\log\left(\frac{q}{2d^2}\right)}\right)$$

steps. Inorder to make sure that we make at least m complete sweeps we let

$$n \geq k \left(\frac{\log(k) + \log(\varepsilon^{-1}) - \log(d)}{\log\left(\frac{q}{2d^2}\right)} + 1\right)$$

and we are done.