The Markov chain convergence theorem: consequences and remaining questions

Given an irreducible and aperiodic Markov chain $(X_0, X_1, ...)$ on a finite state space S we know that there exists exactly one stationary distribution π, and that the distribution of $\mu^{(n)}$ of X_n converges to π as $n \to \infty$.

Known stationary distribution We use the Markov Chain as a tool to simulate a certain distribution π. The MC is constructed so π is the stationary distribution. Simulating X_n for large enough n will give us samples from a distribution close to the stationary one, and theoretically we can test how close π our sampling distribution $\mu^{(n)}$ is.

Unknown stationary distribution We use the markov chain to simulate an asymptotic unknown distribution. Now we cannot check how close to the theoretical distribution out sampling distribution is. This limitation is something we have to consider when anlyzing the result.

The start distribution doesn't matter Eventually the chain "will forget" what distribution it stated in. When using Markov chains as a tool to study some stationary distribution we can just start the chain in any fixed state.

Convergence rate One question remains to be answered. How fast does $d_{\text{TV}}(\mu^{(n)}, \pi)$ approaches zero? When studying a stationary distribution this becomes very important. If we choose n too small then $d_{\text{TV}}(\mu^{(n)}, \pi)$ will bu too large and the distribution we actually are sampling from **may** indeed depend on the initial distribution. A solution to this problem is perfect simulation (see. the course book chapter 10).

Markov chains, a short summary:

	This course	**In general**
State space	Finite $S = \{s_1, ..., s_k\}$.	$\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{Z}^d$, $\{1, ..., q\}^{\mathbb{Z}^d}$, metric spaces.
Existence of a stationary distribution	Requires irreducibility and aperiodicity	Exists for general Markov chains.
Uniqueness of stationary distribution.	Irreducibility and aperiodicity are required	Irreducibility is enough
Convergence towards stationary distribution	Irreducibility and aperiodicity are required.	Irreducibility and aperiodicity are both required but the situation is more complicated. We could have drift towards towards infinity on infinite state spaces.
Reversibility	A reversible distribution is also stationary, but the opposite does not hold in general.	Same situation.

Markov chain Monte Carlo

We now leave the theory of Markov chains and turn to applications. When using Markov chains as algorithmic tool it is important to know when to expect a Markov chain to have a stationary distribution. We first consider two general simulation schemes both generating irreducible Markov chains having reversible distributions.

- The Gibbs sampler
- The Metropolis chain (The Metropolis-Harris chain)

The Gibbs sampler, an example: q-coloring of a graph

Consider a graph $G = (V, E)$ where V is the vertex set and E is the edge set, and let $q \geq 2$. A q-coloring of G is an assignment of colors to each vertex in the color set $\{1, ..., q2\}$ such that no adjecent vertices have the same color.

By a random q-coloring of G we mean a q-coloring chosen uniformly at random from the set of feasible q-colorings, and we write $\rho_{G,q}$ for the corresponding probability distibution on $\{1, ..., q\}^V$. Here we make the assumption that there exists at least one q-colorings.

For a vertex $v \in V$ and an assignment of colors to vertices other than v, the conditional $\rho_{G,q}$-distribution of the color at v is uniform over the set of colors not attained in ξ at some neighbour of v. This follows from a "simple" calculation.

$$\begin{aligned}
\rho_{G,q}(X_{n+1}(v) = a | X_n^v) &= \frac{\rho_{G,q}(\{X_{n+1}(v) = a\} \cap X_n^v)}{\sum_{k=1}^{q} \rho_{G,q}(\{X_{n+1}(v) = k\} \cap X_n^v)} \\
&= \left(\sum_{k=1}^{q} \frac{\rho_{G,q}(\{X_{n+1}(v) = a\} \cap X_n^v)}{\rho_{G,q}(\{X_{n+1}(v) = k\} \cap X_n^v)} \right)^{-1} \\
&= \left(q - \sum_{k=1}^{q} I_{\{X_n(w): w \sim v\}}(k) \right)^{-1}
\end{aligned}$$

where $X_n^v \in \{1, ..., q\}^{V \setminus \{v\}}$ is the configuration on $V \setminus \{v\}$. This is just the uniform distribution over the set of colors not used by any neighbour of v. Given this conditional probability we can look at a step in the simulation scheme for the Gibbs sampler.

1. Pick a vertex $v \in V$ uniformly at random.
2. Pick $X_{n+1}(v)$ uniformly at random from the set of colors that are not attained at any neighbour of v.
3. Let $X_{n+1}(w) = X_n(w)$ for any $w \neq v$ in V.

The question is ...

Does this really give us what we want ?

We need the chain to be irreducible and aperiodic to ensure existence of a unique stationary distribution and convergence towards it. But, does $\mu^{(n)}$ converge to $\rho_{G,q}$?

Problem 7.3 Show that the Gibbs sampler for random q-coloring has $\rho_{G,q}$ as stationary distribution, and that it is aperiodic.

Let $G = (V, E)$ be a graph and $\xi \in \{1, ..., q\}^V$ a configuration, that is, an assignment of colors to each vertex. Let

$$\rho_{G,q}(\xi) = \begin{cases} Z_{G,q}^{-1} & , \xi \text{ is feasible} \\ 0 & , \text{otherwise} \end{cases}$$

where $Z_{G,q}$ is the number of feasible configurations (colorings) of G when using colors $1, .., q$.

Let $X_0, X_1, ...)$ be a Markov chain with state space $\{1, ..., q\}^V$ and transition probabilities given by the Gibbs sampler stated above, and let $v \in V$ be an arbitrary but fixed vertex. To show reversibility we must show that

$$\rho_{G,q}(\xi) P_{\xi,\zeta} = \rho_{G,q}(\zeta) P_{\zeta,\xi}$$

holds for any two states

$$\xi, \zeta \in \{1, ..., q\}^V : \ \xi(w) = \zeta(w) \ , \quad w \in V \setminus \{v\}$$

and any two colors a, b feasible for v.

First some notation. Given $\xi \in \{1, ..., q\}^V$ let $\xi(v)$ denote the color at vertex v and let $\xi^v \ \{1, ..., q\}^{V \setminus \{v\}}$ denote the configuration on all vertices except v.

For any $n \geq 1$ we consider the Markov chain elements $X_n, \ X_{n+1}$. These two elements differ in just one vertex, here denoted v, so $X_n(w) = X_{n+1}(w)$ for $w \neq v$. Assume that $X_n(v) = a$ and $X_{n+1}(v) = b$ for any two colors a, b not used in the neighbourhood of v. Let C_v^q be the colors not used in the set of feasible colors for vertex v.

$$\begin{aligned}
\rho_{G,q}(X_n) P_{X_n, X_{n+1}} &= \rho_{G,q}(X_n) \frac{1}{|V|} \rho_{G,q}(X_n(v) = b | X_n^v) \\
&= \frac{1}{Z_{G,q}} \frac{1}{|V|} \frac{1}{N_v^q} \\
&\overset{(1)}{=} \rho_{G,q}(X_{n+1}) \frac{1}{|V|} \rho_{G,q}(X_{n+1}(v) = a | X_{n+1}^v) \\
&= \rho_{G,q}(X_{n+1}) P_{X_{n+1}, X_n}
\end{aligned}$$

The equality (1) is valid since $\rho_{G,q}(\xi) = Z_{G,q}^{-1}$ for any feasible configuration and that $\rho_{G,q}(\xi(v) = b | \xi^v) = N_v^q)^{-1}$ for any feasible color a at v and any configuration ξ. This is the requirement for reversibility of $\rho_{G,q}$ and we are done!

There is nothing in the transition mechanism forcing us to change the color at vertex v, we are required to update the color, but is could happen that the color is left unchanged. As a consequence we have for any state ξ that

$$\rho_{G,q}(X_{n+1} = \xi | X_n = \xi) > 0$$

making ξ an apriodic state. The same holds for all states making the whole chain apriodic. Note that apriodicity follows without the assumption of irreducibility, since all states have "loops" in the transition graph not just one.

Gibbs sampling in general

The Gibbs sampler is suitable for simulating random variables taking values in some space S^V where both S and V are finite sets. We can think of S^V as the set of all assignments of values in S to each memeber of V. Given some distribution π in S^V we can use the Gibbs sampler to generate samples from π. Given an element $\eta \in S^V$ we denote the configuration of all elements in V except v by η^v. Suppose we have $\xi \in S^V$ and want to update it to ξ' by using the Gibbs sampler.

1. Pick an element $v \in V$ uniformly at random.
2. Condition on the configuration on every element $b \neq a$ and assign $\xi(v)$ value $a \in S$ with probability $\pi(\xi'(v) = a|\xi^v)$.
3. Let $\xi'(w) = \xi(w)$ for any $w \neq v$ in V.

This transition mechanism makes π a reversible distribution.

The Metropolis chain

The Metropolis chain does not use conditional probabilities when updating, instead we are quite free to define between which states we can have a transition. Given the structure of the transition graph the Metropolis chain prescribes specific transition probabilities. To construct a Metropolis chain we need ...

- A state space, for example S^V whenever S and V are finite sets.
- A distribution π on S^V from which we want to generate samples.

... and given this we construct our chain as follows.

The transition graph We decide between which states we allow transitions, this results is a directed graph G with S^V as vertex set. G must be fully connected (to ensure irreducibility for the Metropolis chain) and no vertex in S^V should have too many neighbours, otherwise the computation becomes to heavy.

Transition probabilities The Metropolis chain prescribes the following transition probabilities

$$P_{i,j} = \begin{cases} \frac{1}{d_i} \min\left\{\frac{\pi_j d_i}{\pi_i d_j}, 1\right\} & \text{, if } s_i \text{ and } s_j \text{ are neighbours.} \\ 0 & \text{, if } s_i \neq s_j \text{ are not neighbours.} \\ 1 - \frac{1}{d_i} \sum_{\{l : s_i \sim s_l\}} \min\left\{\frac{\pi_l d_i}{\pi_i d_l}, 1\right\} & \text{, if } s_i = s_j \end{cases}$$

It a easy calculation (see the course book page 51 and 52) to show that π is reversible and thus stationary for P.