Markov chains: convergence

Janko Gravner

MAT 235B

March 16, 2024

Convergence of averages

Let $N_n(y) = \sum_{m=1}^{n} 1_{\{X_m = y\}}$ be the number of visits to y in time interval $[1, n]$.

Theorem

Assume that y is recurrent. Then, for every $x \in S$, P_x-a.s.,

$$\frac{1}{n} N_n(y) \to \frac{1}{E_y T_y} 1_{\{T_y < \infty\}}, \text{ as } n \to \infty.$$

Convergence of averages

Proof.

Assume that we start at y. Recall the successive visit times to y:

$$T_y^0 = 0,$$
$$T_y^k = \inf\{n > T_y^{k-1} : X_n = y\} = \inf\{n \geq 1 : N_n(y) = k\}.$$

Let $\tau_y^k = T_y^k - T_y^{k-1}$, $k = 1, 2, \ldots$. These are i.i.d., by SMP, so by SSLN,

$$\frac{T_y^k}{k} \to E_y T_y, \quad \text{as } k \to \infty,$$

which includes the case when $E_y T_y = \infty$. Note that $T_y^{N_n(y)}$ is the time of the last visit to y in $[0, n]$ and $T_y^{N_n(y)+1}$ is the time of the next visit.

Convergence of averages

Proof, continued.

So,

$$\frac{T_y^{N_n(y)}}{N_n(y)} \le \frac{n}{N_n(y)} \le \frac{T_y^{N_n(y)+1}}{N_n(y)+1} \cdot \frac{N_n(y)+1}{N_n(y)}$$

If $n \to \infty$, $N_n(y) \to \infty$ and therefore

$$\frac{n}{N_n(y)} \to E_y T_y,$$

P_y-a.s.
Now start at x. On $\{T_y = \infty\}$, $N_n(y) = 0$ for all n. Conditioned on $\{T_y < \infty\}$, $\tau_2, \tau_3, \ldots$ are i.i.d., so we can repeat the above argument for P_y. $\square$

Convergence of averages

Corollary

For any Markov chain, and any $x, y \in S$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} p^m(x,y) \to \frac{\rho_{xy}}{E_y T_y}.$$

Proof.

If y is recurrent, we use the previous theorem. By DCT, $\frac{1}{n} E_x N_n(y) \to \frac{\rho_{xy}}{E_y T_y}$, and so

$$\frac{1}{n} \sum_{m=1}^{n} p^m(x,y) \to \frac{\rho_{xy}}{E_y T_y};$$

this is still true if y is transient, because $\sum_{m=1}^{\infty} p^m(x,y) < \infty$ and $P_y(T_y = \infty) > 0$ in that case. $\square$

Weak convergence

Does $\lim_{n\to\infty} p^n(x,y) = \lim_{n\to\infty} P_x(X_n = y)$ exist? The above corollary show that it exists in Cesàro (arithmetic average) sense. The following simple example shows that it does not always exist in the usual sense.

Example. Consiter two-state MC with states 0 and 1 and deterministic transitions $0 \to 1$ and $1 \to 0$. This chain is positive recurrent, with invariant distribution $[1/2, 1/2]$, but $p^n(x,y)$ do not converge because of periodicity.

Let $d_x = \gcd\{n \geq 1 : p^n(x,x) > 0\}$, with $\gcd \emptyset = 0$ be the *period* of $x \in S$. We call the chain *aperiodic* if all states have period 1: $d_x = 1$ for all $x \in S$.

Our final goal is to prove the following two theorems.

Weak convergence

Theorem (Positive recurrent chains)

Suppose that the Markov chain on a countable state space S with transition probability p is irreducible, aperiodic and positive recurrent. Then the chain has a unique invariant probability distribution π on S, and

$$p^n(x, y) = P_x(X_n = y) \to \pi(y)$$

as $n \to \infty$, for all $x, y \in S$.

Weak convergence

Corollary

Suppose that the Markov chain on a countable state space S with transition probability p is irreducible, aperiodic and positive recurrent. For any initial distribution ν on S, $P_\nu(X_n = y) \to \pi(y)$ and so $X_n \xrightarrow{d} \pi$, where π is the unique invariant distribution for the chain.

Proof of Corollary.

We have

$$P_\nu(X_n = y) = \sum_{x \in S} \nu(x) P_x(X_n = y) \to \pi(y),$$

by DCT. $\square$

Weak convergence

Theorem (Chains that are not positive recurrent)

Suppose that the Markov chain on a countable state space S with transition probability p is irreducible, aperiodic and not positive recurrent. Then

$$p^n(x, y) \to 0$$

as $n \to \infty$, for all $x, y \in S$.

In fact, aperiodicity is not necessary in Theorem 2 (but is necessary in Theorem 1). Nevertheless we make that assumption for simplicity.

Observe also that Theorem 2 gives a method to prove positive recurrence, which follows immediately if there is a single pair $x, y \in S$ such that $p^n(x, y)$ does not go to 0.

Finally, use of DCT as in the Corollary implies that in the setting of Theorem 2, $P_\mu(X_n = y) \to 0$ for any initial distribution μ.

Weak convergence

Recall that $d_x = \gcd\{n \geq 1 : p^n(x,x) > 0\}$, with $\gcd \emptyset = 0$, is the period of x, so the chain is aperiodic if and only if $d_x = 1$ for all x. We begin with a lemma from number theory.

Lemma

If $A \subset \mathbb{N} = \{1, 2, \ldots\}$, $\gcd A = 1$, and $m, n \in A$ implies $m + n \in A$, then $\mathbb{N} \setminus A$ is finite.

This lemma is related to the famous *Frobenius coin problem*: given positive integers $a_1 < \cdots < a_n$ with $\gcd(a_1, \ldots, a_n) = 1$, let $g(a_1, \ldots, a_n)$ be the smallest integer not representable as integer linear combination of a_i with *non-negative* coefficients. The lemma implies that $g(a_1, \ldots, a_n) < \infty$, but this number is in general hard to compute. See the book “The Diophantine Frobenius Problem” by Ramírez Alfonsín for more.

Weak convergence

Proof.

First we observe that $D_n = \gcd(A \cap [1, \ldots, n]) = 1$ for a large enough n. (If D_n decreases to some $d > 1$, then all elements of A are divisible by d.) So there exists $\{a_1, \ldots, a_n\} \subset A$, with $\gcd\{a_1, \ldots, a_n\} = 1$. By the Euclidean algorithm, there exist $b_1, \ldots, b_n \in \mathbb{Z}$, such that $\sum_{i=1}^{n} b_i a_i = 1$. Let $b = \max_i |b_i|$ and $a = \sum_i a_i$. If $N \in \mathbb{N}$, write $N = qa + r$, where $q \geq 0$ and $0 \leq r < a$. Then

$$N = q \sum_i a_i + r \sum_i a_i b_i = \sum_i (q + rb_i) a_i.$$

If $N \geq a^2 b$, then $q = \lfloor N/a \rfloor \geq ab$ and so $q + rb_i > ab - ab = 0$. As A is closed under addition, $N \in A$. $\square$

Weak convergence

Lemma (Return Lemma)

If $d_x = 1$, then $\{n \geq 1 : p^n(x,x) > 0\}$ includes all but finitely many integers in $\mathbb{N}$.

Proof.

This follows from the previous lemma, as the set in question is closed under addition. $\square$

Lemma

If $x, y \in S$, and $\rho_{xy} > 0$ and $\rho_{yx} > 0$, then $d_x = d_y$.

Proof.

Choose k, ℓ so that $p^k(x, y) > 0$ and $p^\ell(y, x) > 0$. Then $p^{k+\ell}(y, y) > 0$, so $d_y \mid (k + \ell)$. If $p^n(x, x) > 0$, then $p^{k+n+\ell}(y, y) > 0$ and so $d_y \mid (k + n + \ell)$, and so $d_y \mid n$. This proves that $d_y \mid d_x$ and by symmetry $d_x \mid d_y$. $\square$

Weak convergence

Corollary

Suppose that the Markov chain is irreducible. Then all states have the same period. If $p(x, x) > 0$ *for some* x*, then the chain is aperiodic.*

Weak convergence

To prove Theorem 1, we will use independent *coupling*. Two copies of the chain evolve independently started from possibly different (random or deterministic) initial states. Therefore the state space is now $S \times S$ and the transition probabilities are given by

$$\overline{p}((a,b),(x,y)) = p(a,x)p(b,y).$$

Here is the key lemma.

Lemma (Comparison Lemma)

Assume that p is irreducible, aperiodic, and recurrent. Then the coupled chain is irreducible. Assume also that the coupled chain is recurrent. Then for any two initial distributions μ_1 and μ_2, and any $y \in S$,

$$\sum_{y \in S} |P_{\mu_1}(X_n = y) - P_{\mu_2}(X_n = y)| \to 0$$

as $n \to \infty$.

The lemma states that the total variation distance between the two distributions of X_n goes to 0. We remark that the recurrence of the coupled chain is not necessary; the general case (which we will not need) is proved in the book “Denumerable Markov Chains,” by Kemeny, Snell, and Knapp.

Weak convergence

Proof.

Step 1. The coupled chain (X_n, Y_n) is irreducible. (This is where we use aperiodicity.)

For any $(x_1, y_1), (x_2, y_2) \in S \times S$, we can find an n so that both $p^n(x_1, x_2) > 0$ and $p^n(y_1, y_2) > 0$ (by the Return Lemma). Then

$$\overline{p}^n((x_1, y_1), (x_2, y_2)) = p^n(x_1, x_2) \cdot p^n(y_1, y_2) > 0,$$

proving irreducibility. $\square$

Weak convergence

Proof, continued.

Step 2. Let T be the hitting time of the diagonal for the coupled chain (X_n, Y_n), that is, $D = \{(y, y) : y \in S\}$ and $T = \inf\{n \geq 0 : (X_n, Y_n) \in D\}$. Then the two coordinates have the same distribution after T:

$$P_{(a,b)}(X_n = y, T \leq n) = P_{(a,b)}(Y_n = y, T \leq n),$$

for all $(a, b) \in S \times S$.

Weak convergence

Proof, continued.

To verify this, we divide according to the value of T and the position at time T, then use strong Markov property (as $\{T = k\} \in \mathcal{F}_k$):

$$
\begin{aligned}
&P_{(a,b)}(X_n = y, T \leq n) \\
&= \sum_{k=0}^{n} \sum_{z} P_{(a,b)}(X_n = y, T = k, X_k = Y_k = z) \\
&= \sum_{k=0}^{n} \sum_{z} P_{(z,z)}(X_{n-k} = y) P_{(a,b)}(T = k, X_k = Y_k = z) \\
&= \sum_{k=0}^{n} \sum_{z} P_{(z,z)}(Y_{n-k} = y) P_{(a,b)}(T = k, X_k = Y_k = z) \\
&= P_{(a,b)}(Y_n = y, T \leq n).
\end{aligned}
$$

Weak convergence

Proof, continued.

Step 3. Start the coupled chain at (x, z). Then

$$\begin{aligned} P_x(X_n = y) &= P_{(x,z)}(X_n = y, T \le n) + P_{(x,z)}(X_n = y, T > n) \\ &\le P_z(Y_n = y) + P_{(x,z)}(X_n = y, T > n) \end{aligned}$$

by Step 2, and by symmetry

$$P_z(Y_n = y) \le P_x(X_n = y) + P_{(x,z)}(Y_n = y, T > n).$$

Weak convergence

Proof, continued.

Therefore,

$$
\begin{aligned}
&\sum_y |P_x(X_n = y) - P_z(X_n = y)| \\
&= \sum_y |P_x(X_n = y) - P_z(Y_n = y)| \\
&\le \sum_y \big(P_{(x,z)}(X_n = y, T > n) + P_{(x,z)}(Y_n = y, T > n)\big) \\
&= 2P_{(x,z)}(T > n),
\end{aligned}
$$

which goes to 0 as $n \to \infty$, by irreducibility (Step 1) and recurrence of the coupled chain.

Weak convergence

Proof, continued.

Step 4. Start the coupled chain with $\mu_{X_0} = \mu_1$ and $\mu_{Y_0} = \mu_2$:

$$\begin{aligned}
&\sum_y |P_{\mu_1}(X_n = y) - P_{\mu_2}(X_n = y)| \\
&= \sum_y |P_{\mu_1}(X_n = y) - P_{\mu_2}(Y_n = y)| \\
&\leq \sum_y \sum_{(x,z)} \mu_1(x)\mu_2(z)|P_x(X_n = y) - P_z(Y_n = y)| \\
&= \sum_{(x,z)} \mu_1(x)\mu_2(z) \sum_y |P_x(X_n = y) - P_z(Y_n = y)| \\
&\leq 2 \sum_{(x,z)} \mu_1(x)\mu_2(z) P_{(x,z)}(T > n),
\end{aligned}$$

which goes to 0 by DCT and ends the proof. □

Weak convergence

Lemma (Invariant measure for coupled chain)

If π is an invariant measure for p, then $\overline{\pi}((a, b)) = \pi(a)\pi(b)$ is the invariant measure for $\overline{p}$. Consequently, if the original chain is irreducible, aperiodic and positive recurrent, then so is the coupled chain.

Weak convergence

Proof.

The verification of invariance is routine:

$$\begin{aligned}\sum_{x,y} \overline{\pi}((x,y))\overline{p}((x,y),(a,b)) &= \sum_{x,y} \pi(x)\pi(y)p(x,a)p(y,b) \\ &= \sum_{x} \pi(x)p(x,a) \sum_{y} \pi(y)p(y,b) \\ &= \pi(a)\pi(b).\end{aligned}$$

If the original chain is irreducible and aperiodic, the coupled chain is irreducible. As it has an invariant distribution, the coupled chain is also positive recurrent. For aperiodicity of the coupled chain, observe that $d_{(x,x)}$ divides d_x for all $x \in S$. $\square$

Proof of Theorem 1.

We select the initial states of the two coordinates of the coupled chain as follows: $X_0 = x$, and the distribution of Y_0 is π, that is, $\mu_{Y_0} = \pi$. As the coupled chain is positive recurrent (previous lemma), it is recurrent and Comparison Lemma applies. As $P_\pi(Y_n = y) = \pi(y)$, Comparison Lemma gives that

$$\lim_{n\to\infty} \sum_y |P_x(X_n = y) - \pi(y)| \to 0$$

giving the convergence of μ_{X_n} to π in total variation. (It looks like we proved more, but by Scheffé's theorem weak convergence and convergence in total variation on countable spaces are equivalent.) □

Proof of Theorem 2.

If the chain is transient, we have already proved this; in fact, $\sum_{n=1}^{\infty} p^n(x,y)$ converges in that case. We know that the coupled chain is irreducible. If the coupled chain is transient, the proof is similarly easy, as then for every $x, y \in S$,

$$P_x(X_n = y)^2 = P_{(x,x)}((X_n, Y_n) = (y,y)) \to 0$$

by transience. So we may assume that the coupled chain is recurrent, so that Comparison Lemma applies.

Weak convergence

Proof of Theorem 2, continued.

Fix a $y \in S$. By null recurrence,

$$\sum_{k=0}^{\infty} P_y(T_y > k) = E_y T_y = \infty$$

Fix an $\epsilon > 0$. Choose $K = K(\epsilon)$ so that

$$\sum_{k=0}^{K-1} P_y(T_y > k) > \frac{2}{\epsilon}.$$

Proof of Theorem 2, continued.

For $n \geq K - 1$,

$$
\begin{aligned}
1 &\geq P_x(X_k = y \text{ for some } k = n - K + 1, \dots, n) \\
&= \sum_{k=n-K+1}^{n} P_x(X_k = y, X_m \neq y, m = k + 1, \dots, n) \\
&\qquad\qquad (k \text{ is the last visit to } y \text{ in this time interval}) \\
&= \sum_{k=n-K+1}^{n} P_x(X_k = y) P_y(T_y > n - k) \\
&\qquad\qquad (\text{by the Markov property}) \\
&= \sum_{k=0}^{K-1} P_x(X_{n-k} = y) P_y(T_y > k) \\
&\qquad\qquad (n - k \to k)
\end{aligned}
$$

Weak convergence

Proof of Theorem 2, continued.

Therefore, for every n there exists a $k_n \in \{0, \ldots, K-1\}$ so that $P_x(X_{n-k_n} = y) \leq \epsilon/2$. Now, by Comparison Lemma,

$$P_\mu(X_n = y) - P_\nu(X_n = y) \to 0,$$

as $n \to \infty$, for every initial distributions μ and ν. Choose $\mu = \delta_x$ and ν to be the distribution of X_k when $X_0 = x$. We conclude that, for any fixed k

$$P_x(X_n = y) - P_x(X_{n+k} = y) \to 0.$$

Weak convergence

Proof of Theorem 2, continued.

Therefore, we can choose $n_0 = n_0(\epsilon)$ so that for $n \geq n_0$

$$|P_x(X_{n-k} = y) - P_x(X_n = y)| \leq \epsilon/2,$$

for all $k = 0, \ldots, K-1$. It follows that, for $n \geq n_0$,

$$\begin{aligned} P_x(X_n = y) &\leq P_x(X_{n-k_n} = y) + |P_x(X_{n-k_n} = y) - P(X_n = y)| \\ &\leq \epsilon/2 + \epsilon/2 = \epsilon, \end{aligned}$$

which ends the proof. $\square$