Random Walks: WEEK 3

1 Introduction

Let us first recall an important theorem from last time.

Theorem 1.1 (Stationary distribution)**.** Consider an irreducible Markov chain with transition matrix P. It has a stationary distribution, i.e., a state distribution π^* satisfying $\pi^* = \pi^* P$, if and only if the chain is positive-recurrent.

Let $\pi^{(n)}$ denote the state distribution of the Markov chain at time n. We are interested in the following question: for any given initial distribution $\pi^{(0)}$, does it hold that $\pi^{(n)} \underset{n\to\infty}{\longrightarrow} \pi^*$?

2 The ergodic theorem

Recall that a Markov chain is said to be *aperiodic* if $\mathrm{GCD}\{n : p_{ii}(n) > 0\} = 1$ for any state i.

Definition 2.1 (Ergodicity)**.** A chain is said to be ergodic if it is irreducible, aperiodic and positive-recurrent.

We will prove the following theorem.

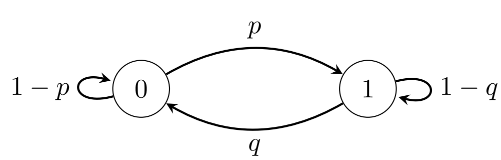
Theorem 2.2 (Ergodic theorem)**.** An ergodic Markov chain admits a unique stationary distribution π^* by Theorem 1.1. This distribution is also a "limiting distribution" in the sense

$$\lim_{n\to\infty} \pi_i^{(n)} = \pi_i^*, \ \forall i \in S$$

Remark 2.3. The state distribution is given by $\pi^{(n)} = \pi^{(0)} P^n$ at any finite time n. The above theorem implies that the limiting distribution does not depend on the initial distribution $\pi^{(0)}$ as $n \to \infty$.

We give an example before starting with the proof.

Example 2.4 (Aperiodicity matters)**.** Consider the following Markov chain with two states $\{0, 1\}$:



It is easy to show that the stationary distribution π^* satisfying

$$\pi^* = \pi^* P = \pi^* \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

has the unique solution $\pi^* = \left(\frac{q}{q+p}, \frac{p}{q+p}\right)$. As a result of Theorem 2.2, this Markov chain has the limiting distribution π^* for any initial distribution if it is ergodic (i.e. irreducible, aperiodic, postive-recurrent). The caveat here is that the assumption on the aperiodicity of the Markov chain is not always satisfied for all p and q.

Suppose $p = q = 1$ and the initial distribution $\pi^{(0)} = (\mathbb{P}\{s = 0\}, \mathbb{P}\{s = 1\}) = (0, 1)$, meaning the chain starts in state 1 with probability one. This Markov chain is not aperiodic. Indeed, we have in this case

$\pi^{(1)} = (1,0)$, $\pi^{(2)} = (0,1)$, $\pi^{(3)} = (1,0)$ and so on. We see that state 1 only has even return times, i.e. $\mathrm{GCD}\{n : p_{11}(n) > 0\} = 2$, thus the chain is periodic with period 2. As a consequence, Theorem 2.2 does not apply to this Markov chain in this case. In fact, one can show that for any initial distribution $\pi^{(0)} = (\alpha, \beta)$ with $\alpha + \beta = 1$, the Markov chain does not converge to the stationary distribution $\left(\frac{1}{2}, \frac{1}{2}\right)$ unless $\alpha = \beta = \frac{1}{2}$.

3 Preliminary tools for the proof

The proof of the ergodic theorem we give here relies on the notions of *total variation distance* and *coupling* that we must first define.

3.1 Total variation distance

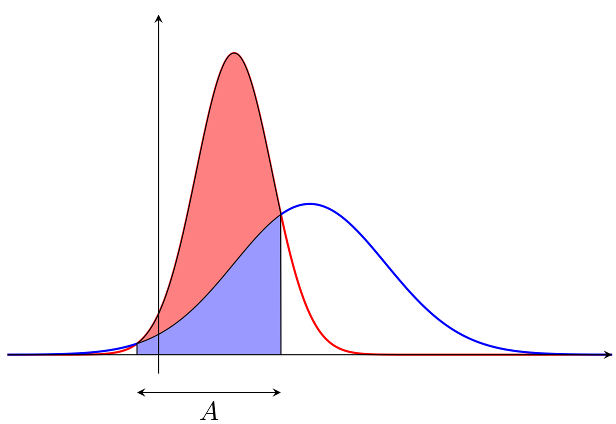
Definition 3.1. Let μ and ν be two probability distributions on the same state space S. The *total variation distance* $d(\mu, \nu)$ is defined as

$$d(\mu, \nu) = \|\mu - \nu\|_{TV} = \sup_{A \subset S} |\mu(A) - \nu(A)|$$

Remark 3.2. The following statements can be deduced from the definition above:

1. $0 \leq \|\mu - \nu\|_{TV} \leq 1$
2. $\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{j \in S} |\mu(j) - \nu(j)|$
3. $d(\mu, \nu)$ is a distance metric, hence it satisfies the symmetry, non-negative and triangle inequalities.

Example 3.3. Consider two random variables $X \sim \mathcal{N}\left(1, \frac{1}{2}\right)$ and $Y \sim \mathcal{N}(2, 1)$. The total variation distance between X and Y is $\|X - Y\|_{TV} \approx 0.458$, which is obtained by choosing $A = [-0.29, 1.62]$.



3.2 Coupling

3.2.1 Coupling of random variables

Definition 3.4. Let μ and ν be two probability distributions over S. A *coupling* between μ and ν is a pair of random variables (X, Y) with joint distribution $\mathbb{P}(X = i, Y = j)$ over $S \times S$ such that the marginals of X and Y are μ and ν. In other words,

$$\mu_i = \mathbb{P}(X = i) \equiv \sum_{j \in S} \mathbb{P}(X = i, Y = j)$$
$$\nu_j = \mathbb{P}(Y = j) \equiv \sum_{i \in S} \mathbb{P}(X = i, Y = j)$$

Example 3.5. We flip a fair coin twice and assign the outcome of each toss to the random variables μ and ν respectively: $\mu_i = \nu_i = \frac{1}{2}$, $i \in \{0, 1\}$.

The two distributions of the random variables (X, Y) below are couplings of μ and ν:

1. $\mathbb{P}\{X = i, Y = j\} = {}^1\!/\!_4$, $(i, j) \in \{0, 1\}^2$
2. $\mathbb{P}\{X = 0, Y = 0\} = \mathbb{P}\{X = 1, Y = 1\} = {}^1\!/\!_2$
   $\mathbb{P}\{X = 1, Y = 0\} = \mathbb{P}\{X = 0, Y = 1\} = 0$

Proposition 3.6 (Coupling and total variation)**.** Let μ and ν be two probability distributions on S. Then

$$\|\mu - \nu\|_{TV} = \inf_{(X,Y)} \mathbb{P}\left\{X \neq Y\right\},$$

where $\inf_{(X,Y)}$ means the infimum on all possible couplings (X, Y) of μ and ν.

In fact we will only need a weaker variant of this proposition, namely

$$\|\mu - \nu\|_{TV} \leq \mathbb{P}\left\{X \neq Y\right\},$$

for any coupling (X, Y) of μ and ν.

Proof. We will prove the weaker bound.

Let A be any subset of S. We have

$$\mu(A) = \mathbb{P}\left(X \in A\right) = \mathbb{P}\left(X \in A, Y \in A\right) + \mathbb{P}\left(X \in A, Y \in A^c\right)$$
$$\nu(A) = \mathbb{P}\left(Y \in A\right) = \mathbb{P}\left(X \in A, Y \in A\right) + \mathbb{P}\left(X \in A^c, Y \in A\right),$$

hence

$$\begin{aligned}\mu(A) - \nu(A) &= \mathbb{P}\left(X \in A, Y \in A^c\right) - \mathbb{P}\left(X \in A^c, Y \in A\right)\\ &\leq \mathbb{P}\left(X \in A, Y \in A^c\right)\\ &\leq \mathbb{P}\left(X \neq Y\right)\end{aligned}$$

By symmetry we also have

$$\nu(A) - \mu(A) \leq \mathbb{P}\left(X \neq Y\right).$$

Thus we get

$$|\mu(A) - \nu(A)| \leq \mathbb{P}\left(X \neq Y\right)$$

Moreover, since we did not impose any particular condition on A, we can take the supremum of the LHS, which gives $\|\mu - \nu\|_{TV} \leq \mathbb{P}\left(X \neq Y\right)$. $\square$

3.2.2 Coupling of Markov chains

Let $(X_n,\, n \geq 0)$ and $(Y_n,\, n \geq 0)$ be two Markov chains on the same state space S having initial distributions $\pi^{(0)} = \mu$ and $\pi^{(0)} = \nu$ respectively. We also assume that both processes have the same transition matrix P (though this condition can be relaxed).

Definition 3.7. We say that the process $((X_n, Y_n),\, n \geq 0)$ on the state space $S \times S$ is a *coupling of Markov chains* $(X_n,\, n \geq 0)$ and $(Y_n,\, n \geq 0)$ if the marginals of (X_n, Y_n) are the processes $(X_n,\, n \geq 0)$ and $(Y_n,\, n \geq 0)$ respectively. In other words,

$$\begin{aligned}
\left(\mu P^n\right)_i &= \mathbb{P}\left(X_n = i\right) \equiv \sum_{j \in S} \mathbb{P}\left(X_n = i, Y_n = j\right) \\
\left(\nu P^n\right)_j &= \mathbb{P}\left(Y_n = j\right) \equiv \sum_{i \in S} \mathbb{P}\left(X_n = i, Y_n = j\right)
\end{aligned}$$

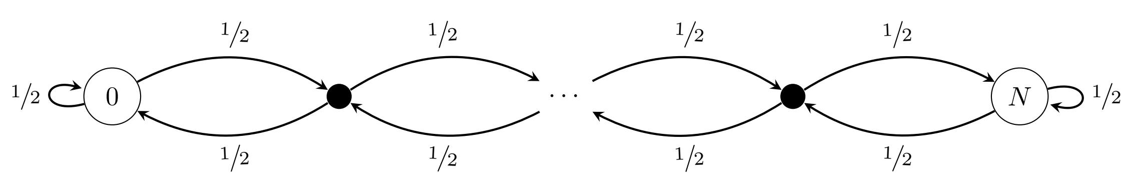
Example 3.8 (Statistical coupling)**.** Consider the processes $(X_n,\, n \geq 0)$ and $(Y_n,\, n \geq 0)$ defined previously. We define the coupling $((X_n, Y_n),\, n \geq 0)$ on the state space $S \times S$ such that the initial distribution is given by $\mu \otimes \nu$ (i.e., $\pi^{(0)}_{i,j} = \mu_i \nu_j$) and the transition matrix is $P \otimes P$ (i.e., $p_{i,j \longrightarrow k,l} = p_{i \to k} p_{j \to l}$).

What does statistical coupling represent? We can picture it as the evolution of two clouds through time in a statistically identical way when starting at two different points.

To verify that this is a valid coupling, we must check that $\left(\mu P^n\right)_i$ and $\left(\nu P^n\right)_j$ are the marginals of $\left((\mu \otimes \nu)\,(P \otimes P)^n\right)_{i,j} = \pi^{(n)}_{i,j}$, $\forall n \geq 0$. This can be proved by induction (the base case $n = 1$ is shown below):

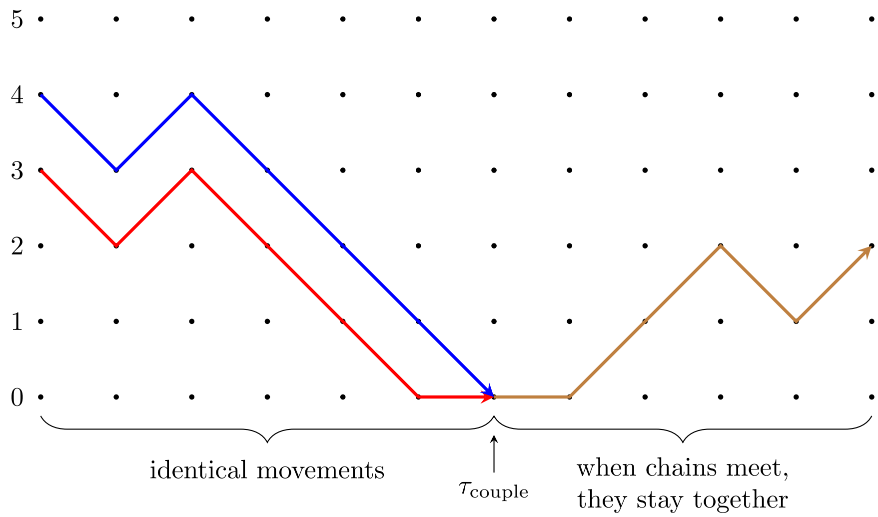
$$\begin{aligned}
\pi^{(1)}_{k,l} &= \sum_{i,j \in S} \pi^{(0)}_{i,j}\, p_{i,j \longrightarrow k,l} = \sum_{i,j \in S} \mu_i \nu_j p_{i \to k} p_{j \to l} = \pi^{(1)}_k \pi^{(1)}_l \\
\sum_{l \in S} \pi^{(1)}_{k,l} &= \pi^{(1)}_k \sum_{l \in S} \pi^{(1)}_l = \pi^{(1)}_k = (\mu P)_k
\end{aligned}$$

Example 3.9 (Grand coupling)**.** Let two random walks $(X_n,\, n \geq 0)$ and $(Y_n,\, n \geq 0)$ be defined on the following state space:



To decide the next state of both processes, we toss a fair coin labeled $\{\pm 1\}$ and move both processes together either one state forward or backward. In the event a process reaches states 0 or N, it takes the self-loop if the coin gave -1 or $+1$ respectively.

If $X_0 = 3$ and $Y_0 = 4$, such a walk would look like the following (with $N = 5$):



We see that at a certain point in time both processes coalesce and remain so forever. We will call this time the *coupling time* τ_{couple}:

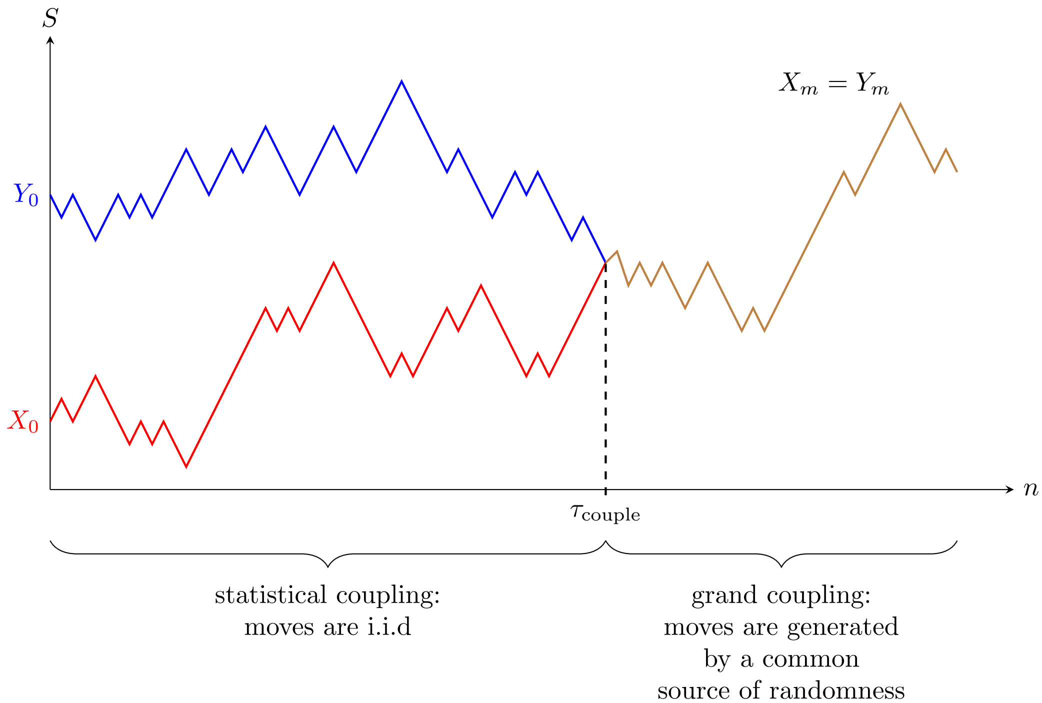
$$\tau_{\text{couple}} = \inf\{n \geq 1 : \, X_n = Y_n\}$$

4 A proof of the ergodic theorem

To prove the ergodic theorem, we use a mixture of the statistical coupling and grand coupling techniques seen previously:

Let $(X_n,\, n \geq 0)$ and $(Y_n,\, n \geq 0)$ be two Markov chains with transition matrices P and initial distributions $X_0 \sim \mu$ and $Y_0 \sim \nu$. We define the coupled process $(Z_n = (X_n, Y_n),\, n \geq 0)$ with the following properties:

1. $Z_0 \sim \mu \otimes \nu$.
2. As long as X_n and Y_n have not coalesced, Z_n's transition matrix is given by $P \otimes P$: Z_n is in a statistical coupling scheme.
3. Once $X_n = Y_n$, we switch to a grand coupling scheme: $X_m = Y_m$, $\forall m \geq n$.



We will see that when X_n and Y_n are ergodic, $\mathbb{P}(\tau_{\text{couple}} < \infty) = 1$.

The following lemmas and corollary will allow us to prove the theorem.

Lemma 4.1. Let X_n, Y_n and Z_n be defined as before. Then

$$\|\mu P^n - \nu P^n\|_{TV} \leq \mathbb{P}(\tau_{\text{couple}} > n)$$

Corollary 4.2. Let $(X_n,\, n \geq 0)$ be an irreducible and positive-recurrent Markov chain. The stationary distribution π^* satisfies for any initial distribution $\pi^{(0)}$

$$\|\pi^* - \pi^{(n)}\|_{TV} \leq \mathbb{P}(\tau_{\text{couple}} > n)$$

Proof. In Lemma 4.1, take $\nu = \pi^{(0)}$ any arbitrary initial distribution and $\mu = \pi^*$ (which exists and is unique by Theorem 1.1). We get

$$\|\mu P^n - \nu P^n\|_{TV} = \|\pi^* P^n - \pi^{(n)}\|_{TV} = \|\pi^* - \pi^{(n)}\|_{TV} \leq \mathbb{P}(\tau_{\text{couple}} > n)$$

□

Lemma 4.3. Let $(X_n,\, n \geq 0)$ and $(Y_n,\, n \geq 0)$ be two ergodic Markov chains. Then

$$\mathbb{P}(\tau_{\text{couple}} > n) \underset{n \to \infty}{\longrightarrow} 0$$

We are now ready to prove the ergodic theorem.

Proof. (Ergodic theorem). The corollary of Lemma 4.1 implied

$$\|\pi^* - \pi^{(n)}\|_{TV} \leq \mathbb{P}(\tau_{\text{couple}} > n)$$

Combining the statement above with Lemma 4.3 gives

$$\lim_{n\to\infty} \|\pi^* - \pi^{(n)}\|_{TV} = 0,$$

which implies the pointwise limit (take the set $A = \{i\}$ in the definition of the total variation distance)

$$\lim_{n\to\infty} \pi_i^{(n)} = \pi_i^*,\ \forall i \in S$$

□

5 Proofs of lemmas

We present the detailed proofs of the previously used lemmas in this section.

5.1 Proof of Lemma 4.1

Consider the two probability distributions μP^n and νP^n at time n. Take the coupling (X_n, Y_n) of μP^n and νP^n: this is a valid coupling because by construction

$$\mathbb{P}(X_n = i) = (\mu P^n)_i$$
$$\mathbb{P}(Y_n = j) = (\nu P^n)_j$$

By Proposition 3.6, we have

$$\|\mu P^n - \nu P^n\|_{TV} \le \mathbb{P}(X_n \neq Y_n)$$

But the event $(X_n \neq Y_n)$ is equivalent to $(\tau_{\text{couple}} > n)$ (by the mixed statistical coupling/grand coupling construction), which proves the lemma.

5.2 Proof of Lemma 4.3

The coupled process $(Z_n = (X_n, Y_n),\, n \ge 0)$ has the following properties:

- **The chain Z is a Markov chain** with the transition probability given by

  $$\begin{aligned} p_{ij\to k\ell} &= \mathbb{P}\{Z_{n+1} = (k,\ell) | Z_n = (i,j)\} \\ &= p_{ik} p_{j\ell} \end{aligned}$$

  where $p_{ik} = \mathbb{P}\{X_{n+1} = k | X_n = i\}$. This fact is easy to verify using the Markovity and independence of X and Y (in the statistical coupling stage).

- **The chain Z is irreducible.**

  Proof. Using the irreducibility and aperiodicity[1] of $(X_n,\, n \ge 0)$, we show in the Appendix that $(X_n,\, n \ge 0)$ satisfies the following property:

  $$\forall i, j \in S, \exists N(i,j) \text{ such that } \forall n \ge N(i,j), p_{ij}(n) > 0$$

  Obviously $(Y_n,\, n \ge 0)$ also satisfies this property. Now for any $i, j, k, \ell \in S$, choose $m > \max\{N(i,j), N(k,\ell)\}$. We will have

  $$\mathbb{P}\{Z_m = (j,\ell) | Z_0 = (i,k)\} = p_{ij}(m)\, p_{k\ell}(m) > 0$$

  □

[1] We remark that this is the only place where aperiodicity is used in the proof of the ergodic theorem.

- **The chain Z is positive-recurrent.**

  Proof. By assumption $(X_n,\, n \geq 0)$ and $(Y_n,\, n \geq 0)$ are irreducible and positive-recurrent, hence by Theorem 1.1 both have stationary distributions π^*. Now define a distribution ν^* on $S \times S$ as

  $$\nu^*_{i,j} = \pi^*_i \pi^*_j \quad (i,j) \in S \times S$$

  Using the fact that π^* is a stationary distribution for $(X_n, n \geq 0)$ and $(Y_n, n \geq 0)$, it is easy to check that $\nu^*_{i,j}$ defined above is indeed a stationary distribution for the chain $(Z_n,\, n \geq 0)$, i.e.

  $$\nu^*_{i,j} = \sum_{k,l \in S} \nu^*_{k,l} p_{k\ell \to ij}$$

  Now use Theorem 1.1 again: since Z is irreducible and has a stationary distribution, it must be positive-recurrent. ☐

Let $Z_0 = (X_0, Y_0) = (i,j)$ and define

$$T_s = \min\{n \geq 1 : Z_n = (s,s)\}$$

for some $s \in S$. In words, this is the first time the trajectories of $(X_n,\, n \geq 0)$ and $(Y_n,\, n \geq 0)$ meet at state s when they start at i and j.

Let m be the smallest time such that $p_{ss\to ij}(m) > 0$. By irreducibility we know that a finite such time exists. This means that the event that the chain goes from (s,s) to (i,j) has non-zero probability. Now, we claim that the following inequality holds:

$$p_{ss\to ij}(m) \cdot (1 - \mathbb{P}(T_s < \infty | Z_0 = (i,j)) \leq 1 - \mathbb{P}(T_s < \infty | Z_n = (s,s))$$

The RHS is the probability that Z leaves (s,s) and never returns; LHS is the probability that Z goes from (s,s) to (i,j) in the least number of steps,[2] then leaves (i,j) but never returns to (s,s). Obviously the event of the LHS is included in the event of the RHS (the event of the LHS implies the event of the RHS). Thus the probability of the LHS is smaller than the probability of the RHS.

The RHS equals zero because Z is recurrent and the first term on the LHS is non-zero because Z is irreducible. This means we must have

$$1 - \mathbb{P}(T_s < \infty | Z_0 = (i,j)) = 0$$

which proves the lemma. ☐

6 Appendix

In the proof of irreducibility of $(Z_n,\, n \geq 0)$ we made use of the following technical statement:

Lemma 6.1. Let $(X_n,\, n \geq 0)$ be irreducible and aperiodic. Then

$$\forall i,j \in S, \exists N(i,j) \text{ such that } \forall n \geq N(i,j), p_{ij}(n) > 0$$

Notice that this is stronger than pure irreducibility because we want $p_{ij}(n) > 0$ for all large enough n (given i,j). This is why aperiodicity is needed. The proof is slightly technical (and has not much to do with probability); but we present it here for completeness.

[2]Here "least number of steps" ensures the walk does not come back to (s,s) when it goes from (s,s) to (i,j).

Proof. For an irreducible aperiodic chain we have for all states $\mathrm{GCD}\{n : p_{jj}(n) > 0\} = 1$. Thus we can find a set of integers $r_1, \ldots, r_k$ such that $p_{jj}(r_k) > 0$ and $\mathrm{GCD}\{r_1, \ldots, r_k\} = 1$.

Claim: for any $r > M$ with M large enough (depending possibly on $r_1, \ldots, r_k$) we can find integers $a_1, \ldots, a_k \in \mathbb{N}$ that are solution of

$$r = a_1 r_1 + \cdots + a_k r_k$$

This claim will be justified at the end of the proof not to disrupt the flow of the main idea.

Since the chain is irreducible, for all i, j we can find some time m such that $p_{ij}(m) > 0$. By the Chapman-Kolmogorov equation we have

$$\begin{aligned} p_{ij}(r+m) &= \sum_{k \in \mathcal{S}} p_{ik}(m) p_{kj}(r) \\ &\geq p_{ij}(m) p_{jj}(r) \end{aligned}$$

Using Chapman-Kolmogorov again and again,

$$\begin{aligned} p_{jj}(r) &= p_{jj}(a_1 r_1 + \cdots + a_k r_k) \\ &= \sum_{\ell_1, \cdots, \ell_k \in \mathcal{S}} p_{j\ell_1}(a_1 r_1) p_{\ell_1 \ell_2}(a_2 r_2) \cdots p_{\ell_k j}(a_k r_k) \\ &\geq p_{jj}(a_1 r_1) p_{jj}(a_2 r_2) \cdots p_{jj}(a_k r_k) \\ &\geq (p_{jj}(r_1))^{a_1} (p_{jj}(r_2))^{a_2} \cdots (p_{jj}(r_k))^{a_k} \end{aligned}$$

We conclude that

$$p_{ij}(r+m) \geq p_{ij}(m) (p_{jj}(r_1))^{a_1} (p_{jj}(r_2))^{a_2} \cdots (p_{jj}(r_k))^{a_k} > 0$$

We have obtained that for all $r > M$ (with M large enough), we have $p_{ij}(r+m) > 0$. Thus we have that $p_{ij}(n) > 0$ *for all* $n > N(i,j) \equiv M + m$. Note that in the above construction M depends on j and m depends on i, j.

It remains to justify the *claim*. For simplicity we do this for $k = 2$. Let $\mathrm{GCD}(a,b) = 1$. We show that for $c > ab$ the equation $c = ax_0 + by_0$ has non-negative solutions (x_0, y_0). If we were allowing negative integers this claim would follow from Bézout's theorem. But here we want non-negative solutions (and maybe that we don't remenber Bézout's theorem anyway?) so we give an explicit argument.

Take $c = ax + by \pmod a$. Then $c \equiv by \pmod a$. Since a and b are coprime, the inverse b^{-1} exists (mod a), so $y \equiv b^{-1}c \pmod a$. Take the smallest integer $y_0 = b^{-1}c \pmod a$ and try now to solve $c = ax + by_0$ for x. Note that $c - by_0 > 0$ because $c > ab$. Note also that $c - by_0$ is divisible by a (since $y_0 - b^{-1}c \equiv 0 \pmod a$). Therefore doing the Euclidean division of $c - by_0$ by a we find x_0 non-negative and satisfying $c - by_0 = ax_0$. We have thus found our solution (x_0, y_0). □