Chapter 11

Markov chains

11.1 Framework for Markov chains

Suppose $S$ is a set with some topological structure that we will use as our state space. Think of $S$ as being $\mathbb{R}^d$ or the positive integers, for example. A sequence of random variables $X_0, X_1, \ldots$, is a Markov chain if

$$
\mathbb{P}(X_{n+1} \in A \mid X_0, \ldots, X_n) = \mathbb{P}(X_{n+1} \in A \mid X_n)
$$

(11.1)

for all $n$ and all measurable sets $A$. The definition of Markov chain has this intuition: to predict the probability that $X_{n+1}$ is in any set, we only need to know where we currently are; how we got there gives no new additional intuition.

Let’s make some additional comments. First of all, we previously considered random variables as mappings from $\Omega$ to $\mathbb{R}$. Now we want to extend our definition by allowing a random variable be a map $X$ from $\Omega$ to $S$, where $(X \in A)$ is $\mathcal{F}$ measurable for all open sets $A$. This agrees with the definition of random variable in the case $S = \mathbb{R}$.

Although there is quite a theory developed for Markov chains with arbitrary state spaces, we will confine our attention to the case where either $S$ is finite, in which case we will usually suppose $S = \{1, 2, \ldots, n\}$, or countable and discrete, in which case we will usually suppose $S$ is the set of positive integers.
We are going to further restrict our attention to Markov chains where
\[ P(X_{n+1} \in A \mid X_n = x) = P(X_1 \in A \mid X_0 = x), \]
that is, where the probabilities do not depend on \( n \). Such Markov chains are said to have stationary transition probabilities.

Define the initial distribution of a Markov chain with stationary transition probabilities by
\[ \mu(i) = P(X_0 = i). \]
Define the transition probabilities by
\[ p(i, j) = P(X_{n+1} = j \mid X_n = i). \]
Since the transition probabilities are stationary, \( p(i, j) \) does not depend on \( n \).

In this case we can use the definition of conditional probability given in undergraduate classes. If \( P(X_n = i) = 0 \) for all \( n \), that means we never visit \( i \) and we could drop the point \( i \) from the state space.

**Proposition 11.1** Let \( X \) be a Markov chain with initial distribution \( \mu \) and transition probabilities \( p(i, j) \). Then
\[ P(X_n = i_n, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i_0) = \mu(i_0)p(i_0, i_1) \cdots p(i_{n-1}, i_n). \] (11.2)

**Proof.** We use induction on \( n \). It is clearly true for \( n = 0 \) by the definition of \( \mu(i) \). Suppose it holds for \( n \); we need to show it holds for \( n + 1 \). For simplicity, we will do the case \( n = 2 \). Then
\[
\begin{align*}
P(X_3 = i_3, X_2 = i_2, X_1 = i_1, X_0 = i_0) &= \mathbb{E}[P(X_3 = i_3 \mid X_0 = i_0, X_1 = i_1, X_2 = i_2); X_2 = i_2, X_1 = i_1, X_0 = i_0] \\
&= \mathbb{E}[P(X_3 = i_3 \mid X_2 = i_2); X_2 = i_2, X_1 = i_1, X_0 = i_0] \\
&= p(i_2, i_3)P(X_2 = i_2, X_1 = i_1, X_0 = i_0).
\end{align*}
\]
Now by the induction hypothesis,
\[ P(X_2 = i_2, X_1 = i_1, X_0 = i_0) = p(i_1, i_2)p(i_0, i_1)\mu(i_0). \] Substituting establishes the claim for \( n = 3 \). \( \square \)

The above proposition says that the law of the Markov chain is determined by the \( \mu(i) \) and \( p(i, j) \). The formula (11.2) also gives a prescription for constructing a Markov chain given the \( \mu(i) \) and \( p(i, j) \).
Proposition 11.2 Suppose \( \mu(i) \) is a sequence of nonnegative numbers with \( \sum_i \mu(i) = 1 \) and for each \( i \) the sequence \( p(i,j) \) is nonnegative and sums to 1. Then there exists a Markov chain with \( \mu(i) \) as its initial distribution and \( p(i,j) \) as the transition probabilities.

Proof. Define \( \Omega = S^\infty \). Let \( \mathcal{F} \) be the \( \sigma \)-fields generated by the collection of sets \( \{(i_0, i_1, \ldots, i_n) : n > 0, i_j \in S\} \). An element \( \omega \) of \( \Omega \) is a sequence \( (i_0, i_1, \ldots) \). Define \( X_j(\omega) = i_j \) if \( \omega = (i_0, i_1, \ldots) \). Define \( \mathbb{P}(X_0 = i_0, \ldots, X_n = i_n) \) by (11.2). Using the Kolmogorov extension theorem, one can show that \( \mathbb{P} \) can be extended to a probability on \( \Omega \).

The above framework is rather abstract, but it is clear that under \( \mathbb{P} \) the sequence \( X_n \) has initial distribution \( \mu(i) \); what we need to show is that \( X_n \) is a Markov chain and that

\[
\mathbb{P}(X_{n+1} = i_{n+1} \mid X_0 = i_0, \ldots, X_n = i_n) = \mathbb{P}(X_{n+1} = i_{n+1} \mid X_n = i_n) = p(i_n, i_{n+1}).
\]

By the definition of conditional probability, the left hand side of (11.3) is

\[
\mathbb{P}(X_{n+1} = i_{n+1} \mid X_0 = i_0, \ldots, X_n = i_n) = \frac{\mathbb{P}(X_{n+1} = i_{n+1}, X_n = i_n, \ldots, X_0 = i_0)}{\mathbb{P}(X_n = i_n, \ldots, X_0 = i_0)} = \frac{\mu(i_0) \cdots p(i_{n-1}, i_n) p(i_n, i_{n+1})}{\mu(i_0) \cdots p(i_{n-1}, i_n)} = p(i_n, i_{n+1})
\]

as desired.

To complete the proof we need to show

\[
\frac{\mathbb{P}(X_{n+1} = i_{n+1}, X_n = i_n)}{\mathbb{P}(X_n = i_n)} = p(i_n, i_{n+1}),
\]

or

\[
\mathbb{P}(X_{n+1} = i_{n+1}, X_n = i_n) = p(i_n, i_{n+1}) \mathbb{P}(X_n = i_n).
\]

(11.5)
CHAPTER 11. MARKOV CHAINS

Now

\[ P(X_n = i_n) = \sum_{i_0, \ldots, i_{n-1}} P(X_n = i_n, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \]

\[ = \sum_{i_0, \ldots, i_{n-1}} \mu(i_0) \cdots p(i_{n-1}, i_n) \]

and similarly

\[ P(X_{n+1} = i_{n+1}, X_n = i_n) \]

\[ = \sum_{i_0, \ldots, i_{n-1}} P(X_{n+1} = i_{n+1}, X_n = i_n, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \]

\[ = p(i_n, i_{n+1}) \sum_{i_0, \ldots, i_{n-1}} \mu(i_0) \cdots p(i_{n-1}, i_n). \]

Equation (11.5) now follows. \(\Box\)

Note in this construction that the \(X_n\) sequence is fixed and does not depend on \(\mu\) or \(p\). Let \(p(i, j)\) be fixed. The probability we constructed above is often denoted \(P^\mu\). If \(\mu\) is point mass at a point \(i\) or \(x\), it is denoted \(P^i\) or \(P^x\). So we have one probability space, one sequence \(X_n\), but a whole family of probabilities \(P^\mu\).

11.2 Examples

Random walk on the integers

We let \(Y_i\) be an i.i.d. sequence of random variables, with \(p = P(Y_i = 1)\) and \(1 - p = P(Y_i = -1)\). Let \(X_n = X_0 + \sum_{i=1}^{n} Y_i\). Then the \(X_n\) can be viewed as a Markov chain with \(p(i, i + 1) = p\), \(p(i, i - 1) = 1 - p\), and \(p(i, j) = 0\) if \(|j - i| \neq 1\). More general random walks on the integers also fit into this framework. To check that the random walk is Markov,

\[ P(X_{n+1} = i_{n+1} \mid X_0 = i_0, \ldots, X_n = i_n) = P(X_{n+1} = i_{n+1} - i_n \mid X_0 = i_0, \ldots, X_n = i_n) \]

\[ = P(X_{n+1} - X_n = i_{n+1} - i_n) \]
using the independence, while
\[
\mathbb{P}(X_{n+1} = i_{n+1} \mid X_n = i_n) = \mathbb{P}(X_{n+1} - X_n = i_{n+1} - i_n \mid X_n = i_n)
= \mathbb{P}(X_{n+1} - X_n = i_{n+1} - i_n).
\]

Random walks on graphs

Suppose we have \(n\) points, and from each point there is some probability of going to another point. For example, suppose there are 5 points and we have \(p(1, 2) = \frac{1}{2}, p(1, 3) = \frac{1}{2}, p(2, 1) = \frac{1}{4}, p(2, 3) = \frac{1}{2}, p(2, 5) = \frac{1}{4}, p(3, 1) = \frac{1}{4}, p(3, 2) = \frac{1}{4}, p(3, 3) = \frac{1}{2}, p(4, 1) = 1, p(5, 1) = \frac{1}{2}, p(5, 5) = \frac{1}{2}\). The \(p(i, j)\) are often arranged into a matrix:
\[
P = \begin{pmatrix}
0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\
\frac{1}{4} & 0 & \frac{1}{2} & 0 & \frac{1}{4} \\
\frac{1}{4} & \frac{1}{4} & \frac{1}{2} & 0 & 0 \\
1 & 0 & 0 & 0 & 0 \\
\frac{1}{2} & 0 & 0 & 0 & \frac{1}{2}
\end{pmatrix}.
\]

Note the rows must sum to 1 since
\[
\sum_{j=1}^{5} p(i, j) = \sum_{j=1}^{5} \mathbb{P}(X_1 = j \mid X_0 = i) = \mathbb{P}(X_1 \in \mathcal{S} \mid X_0 = i) = 1.
\]

Renewal processes

Let \(Y_i\) be i.i.d. with \(\mathbb{P}(Y_i = k) = a_k\) and the \(a_k\) are nonnegative and sum to 1. Let \(T_0 = i_0\) and \(T_n = T_0 + \sum_{i=1}^{n} X_i\). We think of the \(Y_n\) as the lifetime of the \(n\)th light bulb and \(T_n\) the time when the \(n\)th light bulb burns out. (We replace a light bulb as soon as it burns out.) Let
\[
X_n = \min\{m - n : T_i = m \text{ for some } i\}.
\]

So \(X_n\) is the amount of time after time \(n\) until the current light bulb burns out.
If $X_n = j$ and $j > 0$, then $T_i = n + j$ for some $i$ but $T_i$ does not equal $n, n + 1, \ldots, n + j - 1$ for any $i$. So $T_i = (n + 1) + (j - 1)$ for some $i$ and $T_i$ does not equal $(n + 1), (n + 1) + 1, \ldots, (n + 1) + (j - 2)$ for any $i$. Therefore $X_{n+1} = j - 1$. So $p(i, i-1) = 1$ if $i \geq 1$.

If $X_n = 0$, then a light bulb burned out at time $n$ and $X_{n+1}$ is 0 if the next light bulb burned out immediately and $j - 1$ if the light bulb has lifetime $j$. The probability of this is $a_j$. So $p(0, j) = a_{j+1}$. All the other $p(i, j)$’s are 0.

Branching processes

Consider $k$ particles. At the next time interval, some of them die, and some of them split into several particles. The probability that a given particle will split into $j$ particles is given by $a_j$, $j = 0, 1, \ldots$, where the $a_j$ are nonnegative and sum to 1. The behavior of each particle is independent of the behavior of all the other particles. If $X_n$ is the number of particles at time $n$, then $X_n$ is a Markov chain. Let $Y_i$ be i.i.d. random variables with $\mathbb{P}(Y_i = j) = a_j$. The $p(i, j)$ for $X_n$ are somewhat complicated, and can be defined by $p(i, j) = \mathbb{P}(\sum_{m=1}^{i} Y_m = j)$.

Queues

We will discuss briefly the $M/G/1$ queue. The $M$ refers to the fact that the customers arrive according to a Poisson process. So the probability that the number of customers arriving in a time interval of length $t$ is $k$ is given by $e^{-\lambda t} \frac{(\lambda t)^k}{k!}$. The $G$ refers to the fact that the length of time it takes to serve a customer is given by a distribution that is not necessarily exponential. The 1 refers to the fact that there is 1 server.

Suppose the length of time to serve one customer has distribution function $F$ with density $f$. The probability that $k$ customers arrive during the time it takes to serve one customer is

$$a_k = \int_0^\infty e^{-\lambda t} \frac{(\lambda t)^k}{k!} f(t) \, dt.$$ 

Let the $Y_i$ be i.i.d. with $\mathbb{P}(Y_i = k - 1) = a_k$. So $Y_i$ is the number of customers arriving during the time it takes to serve one customer. Let $X_{n+1} = (X_n + Y_{n+1})^+$ be the number of customers waiting. Then $X_n$ is a Markov chain with $p(0, 0) = a_0 + a_1$ and $p(i, i-1+k) = a_k$ if $i \geq 1, k > 1$.

Ehrenfest urns
11.3. MARKOV PROPERTIES

Suppose we have two urns with a total of \( r \) balls, \( k \) in one and \( r - k \) in the other. Pick one of the \( r \) balls at random and move it to the other urn. Let \( X_n \) be the number of balls in the first urn. \( X_n \) is a Markov chain with

\[
p(k, k + 1) = \frac{(r - k)}{r}, \quad p(k, k - 1) = \frac{k}{r}, \quad \text{and} \quad p(i, j) = 0 \quad \text{otherwise}.
\]

One model for this is to consider two containers of air with a thin tube connecting them. Suppose a few molecules of a foreign substance are introduced. Then the number of molecules in the first container is like an Ehrenfest urn. We shall see that all states in this model are recurrent, so infinitely often all the molecules of the foreign substance will be in the first urn. Yet there is a tendency towards equilibrium, so on average there will be about the same number of molecules in each container for all large times.

**Birth and death processes**

Suppose there are \( i \) particles, and the probability of a birth is \( a_i \), the probability of a death is \( b_i \), where \( a_i, b_i \geq 0 \), \( a_i + b_i \leq 1 \). Setting \( X_n \) equal to the number of particles, then \( X_n \) is a Markov chain with \( p(i, i + 1) = a_i \), \( p(i, i - 1) = b_i \), and \( p(i, i) = 1 - a_i - b_i \).

### 11.3 Markov properties

**Proposition 11.3** \( \mathbb{P}^x(X_{n+1} = z \mid X_1, \ldots, X_n) = \mathbb{P}^{X_n}(X_1 = z) \), \( \mathbb{P}^x \)-a.s.

**Proof.** The right hand side is measurable with respect to the \( \sigma \)-field generated by \( X_n \). We therefore need to show that if \( A \) is in \( \sigma(X_n) \), then

\[
\mathbb{P}^x(X_{n+1} = z, A) = \mathbb{E}^x[\mathbb{P}^{X_n}(X_1 = z); A].
\]

\( A \) is of the form \( (X_n \in B) \) for some \( B \subset \mathcal{S} \), so it suffices to show

\[
\mathbb{P}^x(X_{n+1} = z, X_n = y) = \mathbb{E}^x[\mathbb{P}^{X_n}(X_1 = z); X_n = y] \tag{11.6}
\]

and sum over \( y \in B \).

Note the right hand side of (11.6) is equal to

\[
\mathbb{E}^x[\mathbb{P}^y(X_1 = z); X_n = y],
\]
while

$$P^y(X_1 = z) = \sum_i P^y(X_0 = i, X_1 = z) = \sum_i 1_{\{y\}}(i)p(i, z) = p(y, z).$$

Therefore the right hand side of (11.6) is

$$p(y, z)P^x(X_n = y).$$

On the other hand, the left hand side of (11.6) equals

$$P^x(X_{n+1} = z \mid X_n = y)P^x(X_n = y) = p(y, z)P^x(X_n = y).$$

\[\square\]

**Theorem 11.4**

$$P^x(X_{n+1} = i_1, \ldots, X_{n+m} = i_m \mid X_1, \ldots, X_n) = P^x_n(X_1 = i_1, \ldots, X_m = i_m).$$

**Proof.** Note this is equivalent to

$$\mathbb{E}^x[f_1(X_{n+1}) \cdots f_m(X_{n+m}) \mid X_1, \ldots, X_n] = \mathbb{E}^{X_n}[f_1(X_1) \cdots f_m(X_m)].$$

To go one way, we let $f_k = 1_{\{i_k\}}$, to go the other, we multiply the conditional probability result by $f_1(i_1) \cdots f_m(i_m)$ and sum over all possible values of $i_1, \ldots, i_m$.

We'll use induction. The case $m = 1$ is the previous proposition. Let us suppose the result holds for $m$ and prove that it holds for $m + 1$. Let

$$C = (X_{n+1} = i_1, \ldots, X_{n+m-1} = i_{m-1})$$

and

$$D = (X_1 = i_1, \ldots, X_{m-1} = i_{m-1}).$$

Set

$$\varphi(z) = P^z(X_1 = i_{m+1}).$$

Let $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. We then have

$$P^x(X_{n+1} = i_1, \ldots, X_{n+m} = i_m, X_{n+m+1} = i_{m+1} \mid \mathcal{F}_n) = \mathbb{E}^x[P^x(C, X_{n+m} = i_m, X_{n+m+1} = i_{m+1} \mid \mathcal{F}_{n+m+1}) \mid \mathcal{F}_n]$$

$$= \mathbb{E}^x[1_C 1_{\{i_m\}}(X_{n+m})P^x(X_{n+m+1} = i_{m+1} \mid \mathcal{F}_{n+m}) \mid \mathcal{F}_n]$$

$$= \mathbb{E}^x[1_C(\varphi 1_{\{i_m\}})(X_{n+m}) \mid \mathcal{F}_n].$$
11.3. MARKOV PROPERTIES

By the induction hypothesis, this is equal to
\[ E^{X_n}[1_D(\varphi 1_{\{m\}})(X_m)]. \]

For any \( y \),
\[
E^y[1_D(\varphi 1_{\{m\}})(X_m)] \\
= E^y[1_D1_{\{m\}}(X_m)P^{X_m}(X_1 = i_{m+1})] \\
= E^y[1_D1_{\{m\}}(X_m)P^y(X_{m+1} = i_{m+1} \mid F_m)] \\
= E^y[1_D1_{\{m\}}(X_m)1_{\{m+1\}}(X_{m+1})].
\]

\[ \square \]

The strong Markov property is the same as the Markov property, but where the fixed time \( n \) is replaced by a stopping time \( N \).

**Theorem 11.5** If \( N \) is a finite stopping time, then
\[ \mathbb{P}^x(X_{N+1} = i_1 \mid F_N) = \mathbb{P}^{X_N}(X_1 = i_1), \]
and
\[ \mathbb{P}^x(X_{N+1} = i_1, \ldots, X_{N+m} = i_m \mid F_N) = \mathbb{P}^{X_N}(X_1 = i_1, \ldots, X_m = i_m), \]
both \( \mathbb{P}^x \)-a.s.

**Proof.** We will show
\[ \mathbb{P}^x(X_{N+1} = j \mid F_N) = \mathbb{P}^{X_N}(X_1 = j). \]

Once we have this, we can proceed as in the proof of the previous theorem to obtain the second result. To show the above equality, we need to show that if \( B \in F_N \), then
\[ \mathbb{P}^x(X_{N+1} = j, B) = E^x[\mathbb{P}^{X_N}(X_1 = j); B]. \quad (11.7) \]

Recall that since \( B \in F_N \), then \( B \cap (N = k) \in F_k \). We have
\[
\mathbb{P}^x(X_{N+1} = j, B, N = k) = \mathbb{P}^x(X_{k+1} = j, B, N = k) \\
= E^x[\mathbb{P}^x(X_{k+1} = j \mid F_k); B, N = k] \\
= E^x[\mathbb{P}^{X_k}(X_1 = j); B, N = k] \\
= E^x[\mathbb{P}^{X_N}(X_1 = j); B, N = k].
\]
Now sum over $k$; since $N$ is finite, we obtain our desired result. \hfill \Box

We will need the following corollary. We use the notation $T_y = \min\{n > 0 : X_n = y\}$, the first time the Markov chain hits the state $y$.

**Corollary 11.6** Let $U$ be a finite stopping time. Let $V = \min\{n > U : X_n = y\}$, the first time after $U$ that the chain hits $y$. Then

$$
P^x(V < \infty \mid F_U) = P^{X_U}(T_y < \infty).
$$

**Proof.** We can write

$$(V < \infty) = \bigcup_{n=1}^{\infty}(V = U + n)
= \bigcup_{n=1}^{\infty}(X_{U+1} \neq y, \ldots, X_{U+n-1} \neq y, X_{U+n} = y).$$

By the theorem

$$
P^x(X_{U+1} \neq y, \ldots, X_{U+n-1} \neq y, X_{U+n} = y)
= P^{X_U}(X_1 \neq y, \ldots, X_{n-1} \neq y, X_n = y)
= P^{X_U}(T_y = n).
$$

Now sum over $n$. \hfill \Box

Another way of expressing the Markov property is through the Chapman-Kolmogorov equations. Let $p^n(i, j) = P(X_n = j \mid X_0 = i)$.

**Proposition 11.7** For all $i, j, m, n$ we have

$$
p^{n+m}(i, j) = \sum_{k \in S} p^n(i, k)p^m(k, j).
$$

**Proof.** We write

$$
P(X_{n+m} = j, X_0 = i) = \sum_k P(X_{n+m} = j, X_n = k, X_0 = i)
= \sum_k P(X_{n+m} = j \mid X_n = k, X_0 = i)P(X_n = k \mid X_0 = i)P(X_0 = i)
= \sum_k P(X_{n+m} = j \mid X_n = k)p^n(i, k)P(X_0 = i)
= \sum_k p^m(k, j)p^n(i, k)P(X_0 = i).
$$
If we divide both sides by $P(X_0 = i)$, we have our result.

Note the resemblance to matrix multiplication. It is clear if $P$ is the matrix made up of the $p(i, j)$, then $P^n$ will be the matrix whose $(i, j)$ entry is $p^n(i, j)$.

### 11.4 Recurrence and transience

Let

$$T_y = \min\{i > 0 : X_i = y\}.$$ 

This is the first time that $X_i$ hits the point $y$. Even if $X_0 = y$ we would have $T_y > 0$. We let $T_y^k$ be the $k$-th time that the Markov chain hits $y$ and we set

$$r(x, y) = \mathbb{P}^x(T_y < \infty),$$ 

the probability starting at $x$ that the Markov chain ever hits $y$.

**Proposition 11.8** $\mathbb{P}^x(T_y^k < \infty) = r(x, y)r(y, y)^{k-1}$.

**Proof.** The case $k = 1$ is just the definition, so suppose $k > 1$. Let $U = T_y^{k-1}$ and let $V$ be the first time the chain hits $y$ after $U$. Using the corollary to the strong Markov property,

$$\mathbb{P}^x(T_y^k < \infty) = \mathbb{P}^x(V < \infty, T_y^{k-1} < \infty)$$

$$= \mathbb{E}^x[\mathbb{P}^x(V < \infty \mid \mathcal{F}_{T_y^{k-1}}); T_y^{k-1} < \infty]$$

$$= \mathbb{E}^x[\mathbb{P}^x(T_y^{k-1} < \infty); T_y^{k-1} < \infty]$$

$$= \mathbb{E}^x[\mathbb{P}^y(T_y < \infty); T_y^{k-1} < \infty]$$

$$= r(y, y)\mathbb{P}^x(T_y^{k-1} < \infty).$$

We used here the fact that at time $T_y^{k-1}$ the Markov chain must be at the point $y$. Repeating this argument $k - 2$ times yields the result.

We say that $y$ is recurrent if $r(y, y) = 1$; otherwise we say $y$ is transient.

Let

$$N(y) = \sum_{n=1}^{\infty} 1(X_n = y).$$
CHAPTER 11. MARKOV CHAINS

**Proposition 11.9**  
*y* is recurrent if and only if $E^y N(y) = \infty$.

**Proof.** Note

$$E^y N(y) = \sum_{k=1}^{\infty} P^y (N(y) \geq k) = \sum_{k=1}^{\infty} P^y (T^k_y < \infty)$$

$$= \sum_{k=1}^{\infty} r(y, y)^k.$$

We used the fact that $N(y)$ is the number of visits to $y$ and the number of visits being larger than $k$ is the same as the time of the $k$-th visit being finite. Since $r(y, y) \leq 1$, the left hand side will be finite if and only if $r(y, y) < 1$. □

Observe that

$$E^y N(y) = \sum_n P^y (X_n = y) = \sum_n p^n (y, y).$$

If we consider simple symmetric random walk on the integers, then $p^n (0, 0)$ is 0 if $n$ is odd and equal to $\binom{n}{n/2} 2^{-n}$ if $n$ is even. This is because in order to be at 0 after $n$ steps, the walk must have had $n/2$ positive steps and $n/2$ negative steps; the probability of this is given by the binomial distribution. Using Stirling’s approximation, we see that $p^n (0, 0) \sim c / \sqrt{n}$ for $n$ even, which diverges, and so simple random walk in one dimension is recurrent.