Contents

Chapter 1

Introduction and Review of Probability Theory

These notes contain the main definitions and results and a small number of examples. The majority of the examples covered in the module are not in the printed notes but will be provided and worked through in lectures.

1.1 Introduction

Realistic modelling of real world systems such as business, economics, finance, biology, medicine, weather or climate prediction etc. often requires the inclusion of probabilistic elements, i.e., stochastic modelling to deal with uncertainties in the systems (e.g. human decisions in the financial market) or high complexity of the system (e.g., in case of the weather which is a chaotic system).

This module is about *stochastic processes* which are families of random variables X_t, where t is time and X_t lives in some state space to be specified.

1.2 Review of Probability Theory

- (Properties of probabilities) Ω set of events;
  (i) $\Pr(\emptyset) = 0$ (impossible event)
  (ii) $\Pr(\Omega) = 1$ (certain event)
  (iii) $0 \leq \Pr(A) \leq 1$ for all $A \in \Omega$.
  (iv) If $A, B \in \Omega$ are disjoint, i.e., $A \cap B = \emptyset$ then $\Pr(A \cup B) = \Pr(A) + \Pr(B)$.
  (v) let $A = A_1 \cup A_2 \cup \ldots \cup A_n \cup \ldots$ be a disjoint union, i.e., $A_i \cap A_j = \emptyset$ for $i \neq j$. Then $\Pr(A) = \sum_{j=1}^{\infty} \Pr(A_j)$ *(Law of total probability)*.
- Conditional probability: $A, B \subset \Omega$, Ω set of elementary events,

$$\Pr(A \mid B) := \Pr(A \cap B)/\Pr(B).$$

- $A, B \subset \Omega$ independent events if $\Pr(A \mid B) = \Pr(A)$ and $\Pr(B \mid A) = \Pr(B)$.

- **Lemma** $A \subseteq \Omega$, $\Omega = B_1 \cup B_2 \cup \ldots \cup B_n$ with $n \in \mathbb{N} \cup \{\infty\}$, $B_j \cap B_i = \emptyset$ for $i \neq j$. Then

$$\Pr(A) = \sum_{i=1}^{n} \Pr(A \mid B_i)\Pr(B_i).$$

Proof. Let $A_i = A \cap B_i$. Then $A_i \cap A_j = \emptyset$ for $i \neq j$ and $A = A_1 \cup A_2 \cup \ldots \cup A_n \cup \ldots$. So by the law of total probability

$$\Pr(A) = \sum_{i=1}^{n} \Pr(A_i) = \sum_{i=1}^{n} \Pr(A \cap B_i) = \sum_{i=1}^{n} \Pr(A \mid B_i)\Pr(B_i).$$

□

- A random variable (RV) X is a variable that takes its values "by chance", i.e., $X : \Omega \to V$. If V is a finite or denumerable set then X is called discrete RV.

- Let $X : \Omega \to V = \{x_1, x_2, \ldots\}$ be a discrete RV, then $p : V \to [0, 1]$ with $p_i = \Pr(X = x_i)$ is the probability mass function (distribution) of X.

- Let $X : \Omega \to V = \{x_1, x_2, \ldots\}$ be a discrete RV, then

$$E(X) = \sum_{n=1}^{\infty} x_n \Pr(X = x_n)$$

is the expected value of X.

- **Example:** (Binomial distribution) Consider n independent events $A_1, A_2, \ldots, A_n$ with $p = \Pr(A_i)$, $i = 1, \ldots, n$. Let Y be the total number of events that occur. Then

$$p_k = \Pr(Y = k) = \binom{n}{k} p^k (1-p)^{n-k}, \; E(Y) = np.$$

where $\binom{n}{k}$ is the number or possible choices of picking k elements out of n. A concrete example would be to toss a coin n times, with A_i the event that the ith toss results in a "head" and Y the number of "heads" thrown in n tosses.

Chapter 2

Random walks

2.1 Difference equations

Definition 2.1.1. *We call*

$$a_n y_{n+k} + a_{n-1} y_{n+k-1} + \ldots + a_0 y_k = 0 \quad \textit{for all} \quad k \in \mathbb{Z} \quad (HDE)$$

a homogeneous linear difference equation *(HDE) with constant coefficients* $a_0, \ldots, a_n$ *for the sequence* $\{y_n\}$.

Example 2.1.2. *Check that any* $y_k = C_1 2^k + C_2 3^k$ *where* $C_1, C_2 \in \mathbb{R}$ *is a solution for*

$$y_{k+2} - 5y_{k+1} + 6y_k = 0.$$

In order to solve (HDE) substitute $y_k = r^k$ where $r \in \mathbb{R}$. Then (HDE) becomes

$$a_n r^{n+k} + a_{n-1} r^{n+k-1} + \ldots + a_0 r^k = 0.$$

Dividing by r^k assuming $r \neq 0$ gives

$$Q(r) = a_n r^n + a_{n-1} r^{n-1} + \ldots + a_0 = 0,$$

a polynomial of degree n in r. So there are n roots of $Q(r)$ (which may be complex and may not be distinct). We consider two cases:

Case 1: All roots $\{r_1, r_2, \ldots, r_n\}$ of $Q(r) = 0$ are real and distinct. Then $y_k = r_1^k$, $y_k = r_2^k$, …, $y_k = r_n^k$ are solutions of (HDE) and we have the following:

Theorem 2.1.3. *Let* $Q(r) = a_n r^n + a_{n-1} r^{n-1} + \ldots + a_0$. *If all roots* $\{r_1, r_2, \ldots, r_n\}$ *of* $Q(r) = 0$ *are real and distinct, then any solution of (HDE) is of the form*

$$y_k = c_1 r_1^k + c_2 r_2^k + \ldots + c_n r_n^k$$

where $c_1, \ldots, c_n \in \mathbb{R}$.

Case 2: Some roots are equal, say $r_1 = r_2$. Then

$$y_k = (C_1 + C_2 k) r_1^k$$

is a solution of (HDE) for any $C_1, C_2 \in \mathbb{R}$.

Proof. Since $Q(r_1) = 0$ we know that $C_1 r_1^k$ is a solution. Since r_1 is a double root of $Q(r)$ we have

$$Q'(r_1) = 0 = n a_n r_1^{n-1} + (n-1) a_{n-1} r_1^{n-2} + \ldots + a_1.$$

Then

$$\begin{aligned} 0 = r_1^{k+1} Q'(r_1) + k r_1^k Q(r_1) &= n a_n r_1^{n+k} + (n-1) a_{n-1} r_1^{n+k-1} + \ldots + a_1 r_1^{k+1} \\ &+ k a_n r_1^{n+k} + k a_{n-1} r_1^{n+k-1} + \ldots + a_0 k r_1^k \\ &= a_n (n+k) r_1^{n+k} + a_{n-1}(n+k-1) r_1^{n+k-1} + \ldots + a_1(1+k) r_1^{k+1} + a_0 k r_1^k \end{aligned}$$

which proves that $y_k = k r_1^k$ solves (HDE) too. □

More generally we have:

Theorem 2.1.4. *If $r = r_1 = \ldots = r_\ell$ is a root of multiplicity ℓ of $Q(r) = 0$ then*

$$y_k = (c_1 + c_2 k + \ldots + c_\ell k^{\ell-1}) r^k$$

is a solution of (HDE).

Example 2.1.5. *Find the general solution of*

$$y_{k+2} - 6 y_{k+1} + 8 y_k = 0$$

and the particular solution with $y_0 = 3$, $y_1 = 2$.

Remark 2.1.6. *Note that cases 1 and 2 do not cover all cases, but in our applications complex roots will not occur.*

Definition 2.1.7. *We call*

$$a_n y_{n+k} + a_{n-1} y_{n+k-1} + \ldots + a_0 y_k = q_k \quad \textit{for all} \quad k \in \mathbb{Z} \quad (DE)$$

a non-homogeneous linear difference equation *(DE) with constant coefficients $a_0, \ldots, a_n$.*

Theorem 2.1.8. *If y_k^{hom} is a solution of (HDE) and y_k^{part} is a particular solution of (DE) then $y_k = y_k^{\text{hom}} + y_k^{\text{part}}$ is a solution of (DE).*

Example 2.1.9. *Check that $y_k^{\text{part}} = 4k + 6$ is a particular solution of*

$$y_{k+2} - 5 y_{k+1} + 6 y_k = 8k$$

and find the general solution of this DE.

Guidelines 2.1.10. (Finding a particular solution) *To find particular solutions in the case $q_k = \beta^k$ we try $y_k^{\text{part}} = A\beta^k$. For q_k a polynomial of degree m in k,*

$$q_k = c_m k^m + c_{m-1} k^{m-1} + \ldots + c_0$$

we try

$$y_k^{\text{part}} = b_{m+\ell} k^{m+\ell} + b_{m+\ell-1} k^{m+\ell-1} + \ldots b_m k^m + \ldots + b_0$$

as polynomial of degree $m + \ell$ in k where ℓ is the multiplicity of the root $r = 1$ of $Q(r) = 0$.

Example 2.1.11. *Find the general solution of*

$$y_{k+2} - y_k = 4.$$

2.2 Random Walks and Gambler's Ruin

2.2.1 Example: Gambling Problem

Suppose player A starts with $\pounds 3$ and player B with $\pounds 2$. At each bet A wins $\pounds 1$ from B with probability $1/3$ and B wins $\pounds 1$ from A with probability $2/3$. The game continues until either player wins all money. We want to compute

$$\begin{aligned} p_A &= \Pr(\text{A wins all money}), \\ p_B &= \Pr(\text{B wins all money}), \\ p_C &= \Pr(\text{game goes on forever}). \end{aligned}$$

Then

$$p_A + p_B + p_C = 1.$$

Let us concentrate on player A. Let A_n be the amount of money that A has after the nth bet, so $A_n \in \{0, 1, 2, 3, 4, 5\}$ which is the state space of the random variable A_n. This is an example of a random walk. We have

$$A_0 = 3, A_1 \in \{2, 4\}, A_2 \in \{1, 3, 5\}$$

etc. We could draw a tree diagram to compute the probabilities for A_n. But a better method is to generalize to the case when A starts with $\pounds k$ and B with $\pounds 5 - k$, where $0 \le k \le 5$. Let

$$u_k = \Pr(A \text{ wins starting with } \pounds k \mid A_0 = k).$$

We are interested in u_3. We know that $u_0 = 0$ since if A starts with $\pounds 0$ he has already lost; also $u_5 = 1$ since if A starts with all the money $\pounds 5$ he has already won. Moreover, by the law of total probability we have

$$\begin{aligned} u_k &= \Pr(A \text{ wins starting with } \pounds k \mid A \text{ wins first bet}) * \Pr(A \text{ wins first bet}) \\ &+ \Pr(A \text{ wins starting with } \pounds k \mid A \text{ loses first bet}) * \Pr(A \text{ loses first bet}) \\ &= \Pr(A \text{ wins starting with } \pounds k \mid A_1 = k+1) * 1/3 \\ &+ \Pr(A \text{ wins starting with } \pounds k \mid A_1 = k-1) * 2/3 \\ &= \Pr(A \text{ wins starting with } \pounds k \mid A_0 = k+1) * 1/3 \\ &+ \Pr(A \text{ wins starting with } \pounds k \mid A_0 = k-1) * 2/3 \\ &= \tfrac{1}{3} u_{k+1} + \tfrac{2}{3} u_{k-1} \end{aligned}$$

So we get the difference equation

$$u_{k+1} - 3u_k + 2u_{k-1} = 0, \quad u_0 = 0,\ u_5 = 1.$$

Let $u_k = r^k$. Then we need to solve

$$r^2 - 3r + 2 = 0 \;\Leftrightarrow\; (r-2)(r-1) = 0.$$

So the general solution is

$$u_k = c_1 2^k + c_2.$$

From

$$0 = u_0 = c_1 + c_2, \quad 1 = u_5 = 32c_1 + c_2$$

we get

$$c_1 = \tfrac{1}{31}, \ \ c_2 = -\tfrac{1}{31}$$

and so

$$u_k = \tfrac{1}{31}(2^k - 1),$$

and therefore

$$u_1 = \tfrac{1}{31}, \ \ u_2 = \tfrac{3}{31}, \ \ u_3 = \tfrac{7}{31} = p_A, \ \ u_2 = \tfrac{15}{31}.$$

2.2.2 The general gambling problem

Assume in a game the total capital is £N, player A starts with £k and player B with £$(N-k)$, assume that A wins each bet with probability p and B with probability $q = 1 - p$. Let A_n be the amount of money that A has at time n. This is an example of a *random walk* on the integers $k \in \{0, 1, \ldots, N\}$ with absorbing boundaries at $0, N$. Let

$$u_k = \Pr(A \text{ wins starting with } \pounds k \mid A_0 = k).$$

Then $u_0 = 0$ since if A starts with $\pounds 0$ he has already lost. Moreover $u_N = 1$ since if A starts with $\pounds N$ he has already won. If $1 \leq k \leq N - 1$ then

$$\begin{aligned} u_k &= \Pr(A \text{ wins starting with } \pounds k + 1) * \Pr(A \text{ wins first bet}) \\ &\quad + \Pr(A \text{ wins starting with } \pounds k - 1) * \Pr(A \text{ loses first bet}) \\ &= p u_{k+1} + q u_{k-1} \end{aligned}$$

Theorem 2.2.1. *The probability for player A to win solves the difference equation*

$$u_k = p u_{k+1} + q u_{k-1}, \quad u_0 = 0, \ u_N = 1,$$

and has the following solution:

$$u_k = \frac{1 - \left(\frac{q}{p}\right)^k}{1 - \left(\frac{q}{p}\right)^N} \quad \textit{for} \quad p \neq q$$

and

$$u_k = \frac{k}{N} \quad \textit{for} \quad p = q = 1/2.$$

Proof.
Case a): Assume $u_k = r^k$. Then we get

$$r^k = p r^{k+1} + q r^{k-1};$$

dividing by r^{k-1} assuming $r \neq 0$ this gives

$$Q(r) = r - pr^2 - q = 0$$

with solution

$$r = 1, \ \ \tfrac{q}{p}.$$

So the general solution is

$$u_k = c_1 + c_2 \left(\tfrac{q}{p}\right)^k.$$

Now apply the boundary conditions:

$$\begin{aligned} u_0 &= 0 = c_1 + c_2 \\ u_N &= 1 = c_1 + c_2 \left(\frac{q}{p}\right)^N . \end{aligned}$$

This gives

$$c_1 = -c_2, \quad c_1 - c_1 \left(\frac{q}{p}\right)^N = 1$$

and so

$$c_1 = \frac{1}{1 - \left(\frac{q}{p}\right)^N}$$

and

$$u_k = \frac{1 - \left(\frac{q}{p}\right)^k}{1 - \left(\frac{q}{p}\right)^N}$$

as claimed.
Case b). If $q = p$ then $Q(r)$ has a double root $r_1 = r_2 = 1$. So the general solution is, by Theorem 2.1.4,

$$u_k = c_1 + c_2 k.$$

Applying the boundary conditions we get

$$\begin{aligned} 0 &= u_0 = c_1 \\ 1 &= u_1 = c_1 + c_2 N, \end{aligned}$$

and so $c_1 = 0$ and $c_2 = 1/N$, giving $u_k = \frac{k}{N}$ as claimed. □

Proposition 2.2.2. *The probability p_∞ for the game to last forever is 0, i.e. $p_A + p_B = 1$ where p_A (p_B) is the probability that player A (B) wins the game.*

Proof.
Case a): $p = q = \frac{1}{2}$. then by Theorem 2.2.1 we have

$$p_A = k/N, \quad p_B = (N-k)/N$$

and so $p_A + p_B = 1$.
Case b): $p \neq q$. Then by Theorem 2.2.1

$$p_A = \frac{1 - \left(\frac{q}{p}\right)^k}{1 - \left(\frac{q}{p}\right)^N}, \quad p_B = \frac{1 - \left(\frac{\tilde{q}}{\tilde{p}}\right)^{\tilde{k}}}{1 - \left(\frac{\tilde{q}}{\tilde{p}}\right)^N}$$

where

$$\tilde{p} = q, \ \ \tilde{q} = p, \ \ \tilde{k} = N - k,$$

and so

$$p_B = \frac{1-\left(\frac{p}{q}\right)^{N-k}}{1-\left(\frac{p}{q}\right)^{N}}.$$

Let $s = \frac{p}{q}$. Then

$$\begin{aligned} p_\infty &= 1 - p_A - p_B = 1 - \frac{1-s^k}{1-s^N} - \frac{1-\left(\frac{1}{s}\right)^{N-k}}{1-\left(\frac{1}{s}\right)^{N}} \\ &= 1 - \frac{1-s^k}{1-s^N} - \frac{s^N - s^k}{s^N - 1} = 0. \end{aligned}$$

□

Remark 2.2.3. *The gain for player A starting with £k is the random variable G with*

$$\Pr(G = N-k) = u_k, \quad \Pr(G = -k) = 1 - u_k$$

Hence the expected gain of player A is

$$E(G) = (N-k)u_k - k(1-u_k) = Nu_k - k.$$

E.g., if $p = q = \frac{1}{2}$ then $E(G) = N\frac{k}{N} - k = 0$.

Theorem 2.2.4. *Let D_k be the expected duration of the game (number of bets) if $A_0 = k$. Then*

$$D_k = pD_{k+1} + qD_{k-1} + 1, \quad D_0 = 0, \quad D_N = 0,$$

with solution

$$D_k = \frac{1}{p-q}\left(\frac{N\left(1-\left(\frac{q}{p}\right)^k\right)}{1-\left(\frac{q}{p}\right)^N} - k\right) \quad \textit{for} \quad p \neq q$$

and

$$D_k = k(N-k) \quad \textit{for} \quad p = q = 1/2.$$

Proof. Let $\Pr(d_k = n)$ be the probability that the game finishes after n plays. Then

$$D_k = E(d_k) = \sum_{n=1}^{\infty} n\Pr(d_k = n)$$

is the expected duration of the game when A starts with £k.

Let A_n be the amount of money player A has after n plays. By assumption $A_0 = k$. Since in the first step player A may win or lose £1, the law of total probability gives

$$\begin{aligned} \Pr(d_k = n) &= \Pr(d_k = n|A_1 = k+1)\Pr(A_1 = k+1|A_0 = k) \\ &\quad +\Pr(d_k = n|A_1 = k-1)\Pr(A_1 = k-1|A_0 = k) \\ &= \Pr(d_k = n|A_1 = k+1)p + \Pr(d_k = n|A_1 = k-1)q. \end{aligned}$$

If A has won the first trial then after this first trial the event that the game last n trials and A starts with $£k$ becomes the event that game lasts $n-1$ trials and A starts with $£(k+1)$. Hence

$$\Pr(d_k = n | A_1 = k+1) = \Pr(d_{k+1} = n-1).$$

Similarly $\Pr(d_k = n | A_1 = k-1) = \Pr(d_{k-1} = n-1)$. Inserting this into the above equation proves that

$$\Pr(d_k = n) = p\Pr(d_{k+1} = n-1) + q\Pr(d_{k-1} = n-1).$$

Therefore

$$\begin{aligned} D_k &= \sum_{n=1}^{\infty} n\Pr(d_k = n) = \sum_{n=1}^{\infty} n\left(p\Pr(d_{k+1} = n-1) + q\Pr(d_{k-1} = n-1)\right) \\ &= p\sum_{n=1}^{\infty} n\Pr(d_{k+1} = n-1) + q\sum_{n=1}^{\infty} n\Pr(d_{k-1} = n-1) \\ &= p\sum_{m=0}^{\infty} m\Pr(d_{k+1} = m) + p\sum_{m=0}^{\infty} \Pr(d_{k+1} = m) \\ &= +q\sum_{m=0}^{\infty} m\Pr(d_{k-1} = m) + q\sum_{m=0}^{\infty} \Pr(d_{k-1} = m) \\ &= pD_{k+1} + p + qD_{k-1} + q = pD_{k+1} + qD_{k-1} + 1. \end{aligned}$$

Here we used that $\sum_{n=0}^{\infty} \Pr(d_{k\pm1} = n) = 1$ because this is a certain event. So we obtain the nonhomogenous linear difference equation

$$D_k = pD_{k+1} + qD_{k-1} + 1.$$

To solve this difference equation we first have to solve the homogeneous system

$$D_k^{\text{hom}} = pD_{k+1}^{\text{hom}} + qD_{k-1}^{\text{hom}}.$$

Note that we have already solved the same equation for u_k, with different boundary conditions, in the proof of Theorem 2.2.1. Setting $D_k^{\text{hom}} = r^k$ we get $r^k = pr^{k+1} + qr^{k-1}$. Dividing by r^{k-1} assuming $r \neq 0$ we get $Q(r) = pr^2 - r + q = 0$. The roots are $r_1 = 1$ and $r_2 = \frac{q}{p}$.

Case a): $p \neq q$. The general solution of the homogeneous system is

$$D_k^{\text{hom}} = C_1 + C_2\left(\frac{q}{p}\right)^k.$$

Now we need to find a particular solution of the nonhomogeneous system. Since $Q(r) = 0$ has a root 1 of multiplicity $\ell = 1$ and since $q_k = 1$ has degree 0, using Guidelines 2.1.10, we try $D_k^{\text{part}} = b_1 k$ as a particular solution. Inserting this into the nonhomogenous equation we get $b_1 k = pb_1(k+1) + qb_1(k-1) + 1$. Comparing coefficients we obtain $0 = b_1(p-q) + 1$ and so $b_1 = \frac{1}{q-p}$ and $D_k^{\text{part}} = \frac{k}{q-p}$. So the general solution of the nonhomogeneous equation is

$$D_k = \frac{k}{q-p} + C_1 + C_2\left(\frac{q}{p}\right)^k.$$

Boundary conditions: We have $D_0 = D_N = 0$ since the duration of the game is 0 if player A has no or all the money in the game. Setting $D_0 = D_N = 0$ we get

$$0 = C_1 + C_2, \quad 0 = \frac{N}{q-p} + C_1 + C_2\left(\frac{q}{p}\right)^N.$$

Therefore

$$C_1 = -C_2 = N(p-q)^{-1}\left(1 - \left(\frac{q}{p}\right)^N\right)^{-1}$$

and

$$D_k = \frac{1}{p-q}\left(\frac{N\left(1-\left(\frac{q}{p}\right)^k\right)}{1-\left(\frac{q}{p}\right)^N} - k\right).$$

Case b): $p = q = \frac{1}{2}$. The general solution of the homogeneus equation

$$D_k^{\text{hom}} = pD_{k+1}^{\text{hom}} + qD_{k-1}^{\text{hom}}.$$

is (see proof of Theorem 2.2.1)

$$D_k^{\text{hom}} = c_1 + c_2 k.$$

Since $Q(r) = 0$ has a root 1 of multiplicity $\ell = 2$ and since $q_k = 1$ has degree 0, using Guidelines 2.1.10, we try $D_k^{\text{part}} = ck^2$ as a particular solution of the nonhomogeneus equation. Then

$$D_k = \tfrac{1}{2}D_{k-1} + \tfrac{1}{2}D_{k+1} + 1$$

becomes

$$ck^2 = \tfrac{1}{2}c(k-1)^2 + \tfrac{1}{2}c(k+1)^2 + 1$$

which holds true if $c = -1$. So the general solution is

$$D_k = c_1 + c_2 k - k^2.$$

Boundary conditions. From $D_0 = D_N = 0$ we get

$$0 = c_1, \quad 0 = c_1 + c_2 N - N^2$$

and so $c_2 = N$ and $D_k = k(N-k)$. □

Remark 2.2.5. (Gambler's Ruin) If player A plays against an infinitely rich opponent, e.g. a casino or a bank, then $N \to \infty$. Let $s = \frac{q}{p}$.

Case a) $p > q$**.** Here the probability for A to win a bit is bigger than for B so that $s < 1$. In this case, by Theorem 2.2.1,

$$p_A = \lim_{N\to\infty} u_k = \lim_{N\to\infty}\frac{1-s^k}{1-s^N} = 1 - s^k > 0,$$

and the expected duration of the game is, by Theorem 2.2.4,

$$\lim_{N\to\infty} D_k = \lim_{N\to\infty}\frac{1}{p-q}\left(\frac{N(1-s^k)}{1-s^N} - k\right) = \frac{1-s^k}{p-q}\lim_{N\to\infty} N - \frac{k}{p-q} = \infty.$$

Case b) $p < q$**.** Then $s > 1$ and

$$p_A = \lim_{N\to\infty} u_k = \lim_{N\to\infty}\frac{1-s^k}{1-s^N} = 0,$$

and

$$\lim_{N\to\infty} D_k = \lim_{N\to\infty}\frac{1}{p-q}\left(\frac{N(1-s^k)}{1-s^N} - k\right) = \frac{1-s^k}{p-q}\lim_{N\to\infty}\frac{N}{1-s^N} - \frac{k}{p-q} = \frac{k}{q-p}.$$

Case c) $p = q$**.** Then

$$p_A = \lim_{N\to\infty} u_k = \lim_{N\to\infty}\frac{k}{N} = 0$$

so the player is losing, despite the fact that the game is fair! Moreover

$$\lim_{N\to\infty} D_k = \lim_{N\to\infty}(N-k)k = \infty$$

is the expected duration of the game.

Chapter 3

Markov Chains - An Introduction

3.1 Basic Definitions

Definition 3.1.1. *A* stochastic process *is a family of random variables (RVs)* $\{X_t\}_{t\in T}$. *Two cases arise frequently:*

a) $T = \mathbb{N}_0$*: discrete time stochastic process. In this case we write* $\{X_n\}_{n\in\mathbb{N}_0}$.

b) $T = [0, \infty)$*: continuous time stochastic process.*

If $X : T \to S$ *we call* S *the* state space *of the stochastic process* X_t.

Definition 3.1.2. *A* Markov process $\{X_t\}$ *is a stochastic process with the property that given the values of* X_t *the values of* X_s *for* $s > t$ *only depend on* X_t *and not on* X_r *for* $r < t$.

In short a Markov process is a stochastic process without memory.

Definition 3.1.3. *A* Markov chain *(M.C.)* $\{X_n\}$ *is a discrete time Markov process on a finite or countable state space.*
The Markov property for a M.C. reads:

$$\Pr(X_{n+1} = j \mid X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = i_n) = \Pr(X_{n+1} = j \mid X_n = i).$$

Definition 3.1.4. *We call* $\Pr(X_{n+1} = j \mid X_n = i)$ *the* one-step transition probabilities *of the M.C.* $\{X_n\}$ *(for all* $n \in \mathbb{N}_0$, $i, j \in S$.

Definition 3.1.5. *The M.C. has* stationary transition probabilities *if* $p_{i,j} := \Pr(X_{n+1} = j \mid X_n = i)$ *is independent of* n *for all* i, j. *In this case the M.C. is called* homogeneous.

3.2 Transition Probability Matrices

Definition 3.2.1. $\{X_n\}$ *homogeneous M.C. on state space* $S = \{1, 2, \ldots, N\}$. *Then*

$$P = (p_{ij})_{i,j=1,\ldots,N}$$

is called (one-step) transition probability matrix *(TPM).*

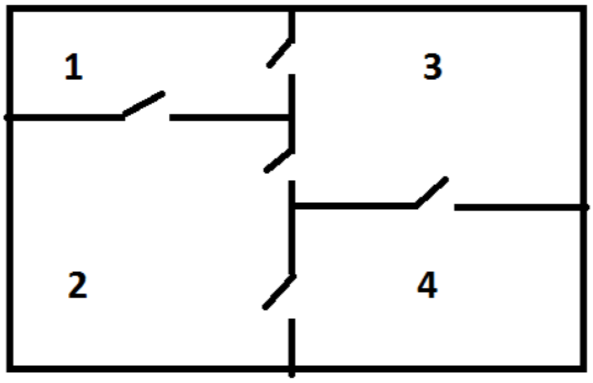

Figure 3.1: Floor plan of house.

Example 3.2.2. Consider a cat in a house with 4 rooms. The doors between rooms are indicated by slashes in the floor plan of Figure 3.1.

If the cat is in one room it chooses one of the doors to another room with equal probability. So it goes from room 1 to room 2 with probability $p_{12} = 1/2$ and to room 3 with probabiity $p_{13} = 1/2$ etc.. What is P?

$$P = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ \vdots & \vdots & \vdots & \vdots \\ p_{41} & p_{42} & p_{43} & p_{44} \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

Example 3.2.3. (Gambler's problem with total capital N) What is the state space and the TPM of this Markov chain?

Definition 3.2.4. *An* (N, N)*-matrix* P *is called* row-stochastic *if*

$$0 \le p_{ij} \le 1, \quad \sum_{j=1}^{N} p_{ij} = 1 \quad \textit{for all} \quad i = 1, \dots, N.$$

Proposition 3.2.5. *The TPM* $P = (p_{ij})_{i,j \in S}$ *of a homogeneous M.C. on* $S = \{1, \dots, N\}$ *is row-stochastic.*

Proof. The entries p_{ij} of P are probabilities, so $p_{ij} \in [0, 1]$. For any state $i \in S$ we have

$$\sum_{j=1}^{N} p_{ij} = 1$$

because the left hand side is the probability that the process has to go to some other state from state i, which is the probability of the certain event. □

Theorem 3.2.6. *Let* $\{X_n\}$ *be a homogeneous M.C. on the state space* $S = \{1, 2, \dots, N\}$ *with TPM* P *and initial distribution* $p_i = \Pr(X_0 = i)$, $i \in S$. *Then*

$$\Pr(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = p_{i_0} p_{i_0 i_1} \dots p_{i_{n-1} i_n}.$$

Proof. We have

$$
\begin{aligned}
&\Pr(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n)\\
&\quad = \Pr(X_n = i_n \mid X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1})\Pr(X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1})\\
&\quad = \Pr(X_n = i_n \mid X_{n-1} = i_{n-1})\Pr(X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1})\\
&\quad = p_{i_{n-1}i_n}\Pr(X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}).
\end{aligned}
$$

Here we used the definition of conditional probability in the first line and the Markov property in the second line. Iterating this argument we get

$$
\begin{aligned}
\Pr(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) &= p_{i_{n-1}i_n}p_{i_{n-2}i_{n-1}}\Pr(X_0 = i_0, X_1 = i_1, \dots, X_{n-2} = i_{n-2})\\
&\ \ \vdots\\
&= p_{i_{n-1}i_n}p_{i_{n-2}i_{n-1}}\cdots p_{i_0 i_1}p_{i_0}.
\end{aligned}
$$

□

Example 3.2.7. We send a binary message through a channel with several stages, each stage has a fixed probability of error α, $0 < \alpha < 1$. Let $X_0 = 0$ and let X_n be the signal received at the nth stage. Assume that $\{X_n\}$ is a homogeneous Markov chain.

a) What is the probability of no error in transmission up to and including stage 2?

b) What is the probability that a correct signal is received at stage 2?

3.3 n-step transition probability matrices

Definition 3.3.1. *Let $\{X_n\}$ be a homogeneous M.C. on the state space $S = \{1, 2, \dots, N\}$. Then $P^{(n)} = (p_{ij}^{(n)})_{i,j\in S}$ with $p_{ij}^{(n)} = \Pr(X_{m+n} = j \mid X_m = i)$ is called the n-*step transition probability matrix *of the Markov chain.*

Example 3.3.2. Consider a Markov chain on the states $\{0, 1, 2\}$ with TPM

$$
P = \begin{pmatrix} 0.5 & 0 & 0.5 \\ 0.1 & 0.9 & 0 \\ 0.2 & 0.6 & 0.2 \end{pmatrix}
$$

What is $p_{00}^{(2)}$?

Theorem 3.3.3. *Let $\{X_n\}$ be a homogeneous M.C. on the state space $S = \{1, 2, \dots, N\}$ with TPM P. Then $P^{(n)} = P^n$. Componentwise*

$$
\forall i, j \quad p_{ij}^{(n)} = \sum_{k=1}^{N} p_{ik}^{(n-1)}p_{kj} \quad \Leftrightarrow \quad P^{(n)} = P^{(n-1)}P, \tag{3.1}
$$

where we define $p_{ij}^{(0)} = 1$ if $i = j$ and $p_{ij}^{(0)} = 0$ if $i \neq j$.

In words (3.1) can be explained as follows: to get from i to j in n steps (the probability of which is $p_{ij}^{(n)}$) the process has to go from state i to some state k in $(n-1)$ steps (the probability of which is $p_{ik}^{(n-1)}$) and then from state k to state j in one step (the probability of which is p_{kj}). Summing over all states k gives (3.1).

For a detailed proof of this theorem we need the following lemma:

Lemma 3.3.4. *If $\{C_k\}_{k=1,\ldots,N}$ are disjoint events such that $\sum_{k=1}^{N} \Pr(C_k) = 1$ then*

$$\Pr(A \mid B) = \sum_{k=1}^{N} \Pr(A \mid B \cap C_k) \Pr(C_k \mid B)$$

Proof of Lemma 3.3.4.

$$\begin{aligned} \Pr(A \mid B) &= \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{1}{\Pr(B)} \sum_{k=1}^{N} \Pr((A \cap B) \cap C_k) \\ &= \sum_{k=1}^{N} \Pr(A \mid B \cap C_k) \Pr(B \cap C_k)/\Pr(B) = \sum_{k=1}^{N} \Pr(A \mid B \cap C_k) \Pr(C_k \mid B). \end{aligned}$$

Proof of Theorem 3.3.3. Now with $A = \{X_n = j\}$, $B = \{X_0 = i\}$ and $C_k = \{X_{n-1} = k\}$ we get

$$\begin{aligned} p_{ij}^{(n)} &= \Pr(X_n = j \mid X_0 = i) = \sum_{k=1}^{N} \Pr(X_n = j \mid X_0 = i, X_{n-1} = k) \Pr(X_{n-1} = k \mid X_0 = i) \\ &= \sum_{k=1}^{N} \Pr(X_n = j \mid X_{n-1} = k) \Pr(X_{n-1} = k \mid X_0 = i) = \sum_{k=1}^{N} p_{kj} p_{ik}^{(n-1)} \\ &= \sum_{k=1}^{N} p_{ik}^{(n-1)} p_{kj} = (P^{(n-1)} P)_{ij}. \quad (*) \end{aligned}$$

In words: To get from i to j in n steps, the probability of which is $p_{ij}^{(n)}$, the process has to go from i to some state k in $(n-1)$ steps, the probability of which is $p_{ik}^{(n-1)}$, and then from state k to state j in one step, the probability of which is p_{kj}. Summing over all states k gives (*).

Hence $P^{(n)} = P^{(n-1)} P$. So if it is true that $P^{(n-1)} = P^{n-1}$ then the above proves that also $P^{(n)} = P^{(n-1)} P = P^{n-1} P = P^n$. For $n = 1$ we know that $P^{(1)} = P^1 = P$. So by induction, $P^{(n)} = P^n$ for all n. □

Example 3.3.5. A particle moves among the states $\{0, 1, 2\}$ with TPM

$$P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}.$$

Compute $p_{00}^{(2)}$ and $p_{00}^{(3)}$.

Corollary 3.3.6. *Let $\{X_n\}$ be a homogeneous Markov chain with TPM P, state space $S = \{1, 2, \ldots, N\}$ and initial distribution $p = (p_i)_{i \in S} = (p_1, \ldots, p_N)$. Then for all $i \in S$,*

$$\Pr(X_n = i) = (pP^n)_i.$$

Proof By Theorem 3.3.3

$$\Pr(X_n = i) = \sum_{k\in S} \Pr(X_n = i \mid X_0 = k) \Pr(X_0 = k) = \sum_{k\in S} p_{ki}^{(n)} p_k = (pP^n)_i.$$

□

Note that p and pP are row vectors.

3.4 First Step Analysis of Markov Chains

First step analysis analyzes the probabilities arising in the first transition for all initial states. We have used first step analysis already for the gambling problem in Section 2.2.2 and will now introduce first step analysis for general Markov chains.

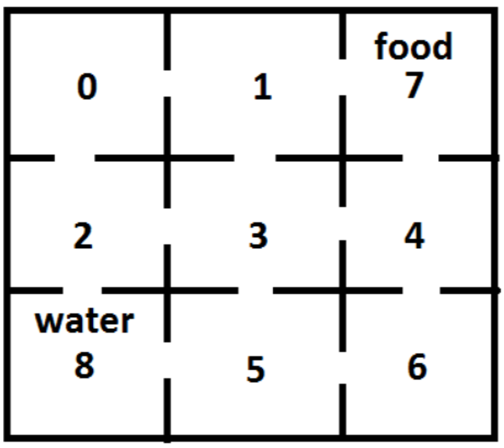

Figure 3.2: Floor plan of house.

Example 3.4.1. Assume a cat is in a house with 9 rooms, see Figure 3.2. If it has k choices to leave a room the cat chooses each room with probability $\frac{1}{k}$. What is the probability that the cat encounters the food before the water given that it starts in room 4? Let u_k be the probability of absorption into the food compartment given that the cat starts in room k. Then by the law of total probability

$$\begin{aligned}
u_0 &= \frac{1}{2}u_1 + \frac{1}{2}u_2 \\
u_1 &= \frac{1}{3}u_0 + \frac{1}{3}u_3 + \frac{1}{3} \\
u_3 &= \frac{1}{4}u_1 + \frac{1}{4}u_2 + \frac{1}{4}u_4 + \frac{1}{4}u_5 \\
u_4 &= \frac{1}{3} + \frac{1}{3}u_3 + \frac{1}{3}u_6 \\
u_5 &= \frac{1}{3}u_3 + \frac{1}{3}u_6 \\
u_6 &= \frac{1}{2}u_4 + \frac{1}{2}u_5 \\
u_7 &= 1 \\
u_8 &= 0.
\end{aligned}$$

Symmetry implies

$$u_0 = u_6, \;\; u_2 = u_5, \;\; u_1 = u_4, \;\; u_3 = \frac{1}{2}.$$

Then we get $u_0 = \frac{1}{2} = u_6$, $u_1 = \frac{2}{3} = u_4$, $u_2 = \frac{1}{3} = u_5$.

Example 3.4.2. Consider the following model of the life span of females in a population. The state space is $S = \{0, 1, \ldots, 5\}$ with state 0 = prepuberty, state 1 = single, state 2 = married, state 3 = divorced, state 4 = widowed, state 5= dead. Suppose the model has the following TPM:

$$P = \begin{pmatrix} 0 & 0.9 & 0 & 0 & 0 & 0.1 \\ 0 & 0.5 & 0.4 & 0 & 0 & 0.1 \\ 0 & 0 & 0.6 & 0.2 & 0.1 & 0.1 \\ 0 & 0 & 0.4 & 0.5 & 0 & 0.1 \\ 0 & 0 & 0.4 & 0 & 0.5 & 0.1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

What is the expected duration in state 2 (married)?

Example 3.4.3. On the state space $S = \{0, 1, 2, 3\}$ consider a M.C. with TPM

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0.1 & 0.6 & 0.1 & 0.2 \\ 0.2 & 0.3 & 0.4 & 0.1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

a) Starting in state 1 what is the probability of absorption in state 0?

b) What is the expected time to absorption (i.e., M.C. ends up in one of the absorbing states) starting from state i.

More generally, let $\{X_n\}$ homogeneous M.C. on the state space $S = \{0, 1, 2, \ldots, N\}$ with TPM P. We label the states such that $\{0, 1, \ldots, r-1\}$ are not absorbing, $\{r, r+1, \ldots, N\}$ are absorbing, i.e. $p_{ii} = 1$ for $i > r$. Then the TPM P of $\{X_n\}$ takes the form

$$P = \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix}$$

where Q is an (r, r)-matrix describing the transitions between non-absorbing states, $q_{ij} = p_{ij}$, I is the $(N+1-r, N+1-r)$ identity matrix, and R is the matrix that describes the transitions from non-absorbing to absorbing states.

We want to compute the probability of absorption into an absorbing state k, the expected number of visits to a state j before absorption and the expected time before absorption starting from state i.

Theorem 3.4.4. *Let k be an absorbing state, and let u_i be the probability of being absorbed into state k starting from state i, $i = 0, \ldots, r-1$. Then*

$$u_i = p_{ik} + \sum_{j=0}^{r-1} p_{ij} u_j, \quad i \in \{0, 1, \ldots, r-1\},$$

and $u_j = 0$ for $j \neq k$, $j \geq r$.

Proof. From state i the process can go in one step directly into state k, with probability p_{ik} or it can go to another non-absorbing state $0 \leq i < r$ or into another absorbing state j with $r \leq j \leq N$, with $j \neq k$. Hence

$$\begin{aligned} u_i &= \sum_{j=0}^{N} \Pr(\text{absorption in state } k \mid X_0 = i, X_1 = j) p_{ij} \\ &= p_{ik} + 0 + \sum_{j=0}^{r-1} \Pr(\text{absorption in state } k \mid X_0 = i, X_1 = j) p_{ij} \end{aligned}$$

where we get the 0 term if the process goes in the first step to another absorbing state $j \neq k$ and the sum corresponds to transitions to another non-absorbing state j, $0 \leq j < r$, in the first step. By the Markov property we get

$$u_i = p_{ik} + \sum_{j=0}^{r-1} \Pr(\text{absorption in state } k \mid X_1 = j) p_{ij} = p_{ik} + \sum_{j=0}^{r-1} u_j p_{ij}.$$

□

Let $d_i = \min\{n \geq 0, X_n \geq r \mid X_0 = i\}$ is the time which the process spends in non-absorbing states before being absorbed, starting in state i. Assume $\Pr(d_i < \infty) = 1$ for all i. More generally, let $g : \{0, 1, \ldots, r-1\} \to \mathbb{R}$ and set $g(k) = 0$ for $k \geq r$ (for all absorbing states). We wish to calculate the expected sum of values of g during the stochastic process starting from state i until absorption,

$$w_i = E(\sum_{n=0}^{d_i - 1} g(X_n) \mid X_0 = i).$$

Examples 3.4.5.

a) If $g(i) \equiv 1$ for all non-absorbing states i then $w_i = D_i$ is the expected time until absorption;

b) If $g \equiv 1_{\{j\}}$, where j is a fixed non-absorbing state, i.e., $g(i) = 1$ for $i = j$, and $g(i) = 0$ otherwise, then w_i is the expected number of visits to state j.

Theorem 3.4.6. *With the above notation we have*

$$w_i = g(i) + \sum_{k=0}^{r-1} p_{ik} w_k. \tag{3.2}$$

Proof. Using First Step Analysis, if the process starts in state i then $\sum_{n=0}^{d_i-1} g(X_n)$ always contains $g(X_0) = g(i)$. The expected value of $\sum_{n=1}^{d_i-1} g(X_n)$ starting from state i going to state k in the first step is $p_{ik}w_k$. Summing over all states k we get (3.2).

In more detail:

$$\begin{aligned} w_i &= E(\sum_{n=0}^{d_i-1} g(X_n) \mid X_0 = i) \\ &= E(g(X_0) \mid X_0 = i) + E(\sum_{n=1}^{d_i-1} g(X_n) \mid X_0 = i) \\ &= g(i) + E(\sum_{n=1}^{d_i} g(X_n) \mid X_0 = i) \\ &= g(i) + \sum_{k=0}^{r-1} E(\sum_{n=1}^{d_i} g(X_n) \mid X_1 = k, X_0 = i)\Pr(X_1 = k \mid X_0 = i). \end{aligned}$$

Here we used in the third line that $g(X_{d_i}) = 0$ since X_{d_i} is an absorbing state, and in the last line we used the law of total probability. By the Markov property

$$w_i = g(i) + \sum_{k=0}^{r-1} E(\sum_{n=1}^{d_i} g(X_n) \mid X_1 = k) p_{ik}.$$

Shifting time $n = m + 1$ we get

$$w_i = g(i) + \sum_{k=0}^{r-1} E(\sum_{m=0}^{d_i-1} g(X_m) \mid X_0 = k) p_{ik} = g(i) + \sum_{k=0}^{r-1} w_k p_{ik}.$$

□

Examples 3.4.7.

a) If $w_i = D_i$ is the expected time until absorption starting from state i then for all $i \in \{0, 1, \dots, r-1\}$

$$D_i = 1 + \sum_{k=0}^{r-1} p_{ik} D_k.$$

b) If w_i is the expected number of visits to a given non-absorbing state j starting from state i then

$$w_j = 1 + \sum_{k=0}^{r-1} p_{jk} w_k, \quad w_i = \sum_{k=0}^{r-1} p_{ik} w_k \quad i \neq j.$$

Example 3.4.8. Consider a M.C. on $\{0, 1, 2, 3\}$ with TPM

$$P = \begin{pmatrix} 0.2 & 0.8 & 0 & 0 \\ 0.1 & 0.3 & 0.3 & 0.3 \\ 0.6 & 0.1 & 0 & 0.3 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

a) What is the expected number of visits to state 2 before absorption into state 1?

b) What is the expected number of visits to state 1 before absorption into state 2?

Chapter 4

Long-time Behaviour of Markov Chains

4.1 Transience, Recurrence, and Periodicity

Definition 4.1.1. *Let j be a state in a M.C. with TPM P. Then we call any $n \in \mathbb{N}$ such that $p_{jj}^{(n)} > 0$ a* possible return time *for state j; moreover we call the greatest common divisor of all possible return times for state j the* period $\tau(j)$ *of state j:*

$$\tau(j) = \gcd(n \in \mathbb{N}, p_{jj}^{(n)} > 0).$$

If $p_{jj}^{(n)} = 0$ for all n then we set $\tau = 0$.
If $\tau(j) > 1$ we call the state j periodic*; if $\tau(j) = 1$ we call state j* aperiodic*. A M.C. is called aperiodic if all its states are aperiodic.*

Note that all possible return times for state j are multiples of the period $\tau(j)$ of state j.

Example 4.1.2. Consider a random walk on the integers such that

$$\Pr(X_n = j+1 \mid X_{n-1} = j) = p, \quad \Pr(X_n = j-1 \mid X_{n-1} = j) = q, \quad p+q = 1.$$

To determine the period τ of each state we need to find τ such that the return times are multiples of τ only. Since $p_{jj}^{(n)} = 0$ for n odd the period of each state j is $\tau = 2$.

Example 4.1.3. Consider a random walk on the integers such that

$$\Pr(X_n = j-1 \mid X_n = j) = \alpha, \quad \Pr(X_n = j+1 \mid X_n = j) = \beta, \quad \Pr(X_n = j \mid X_n = j) = \gamma,$$

where

$$\alpha, \beta, \gamma > 0, \quad \alpha + \beta + \gamma = 1.$$

This is an aperiodic Markov chain.

Example 4.1.4. Suppose a Markov chain has state space $\{0, 1, 2, 3\}$ and TPM

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

Then all states have period $\tau = 4$.

Definition 4.1.5. *Let i, j be states of a homogeneous M.C.. Then we denote by $f_{ij}^{(n)}$ the probability that starting at state i the M.C. first reaches state j at time n. We set $f_{ij}^{(0)} = 0$ for $i \neq j$.*

NOTE: in contrast $p_{ij}^{(n)}$ the probability that starting at state i the M.C. is in state j at time n.

Proposition 4.1.6. *Let i, j be states of a homogeneous M.C.. Set $f_{ij} := \sum_{n=1}^{\infty} f_{ij}^{(n)}$. For $i \neq j$ this is the probability that the M.C. ever reaches state j starting from state i; moreover f_{ii} is the probability of ever returning to state i.*

Proof. Let A_n be the event that the M.C. is in state j for the first time at time n starting from state $i \neq j$ and the event that the M.C. first returns to state i at time n starting from state i if $i = j$. Then $A_n \cap A_m = \emptyset$ for $n \neq m$. By definition $f_{ij}^{(n)} = \Pr(A_n)$. Furthermore $\bigcup_{n=1}^{\infty} A_n$ is the event that there is some time n such that the M.C. is in state j. Hence

$$\Pr(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \Pr(A_n) = \sum_{n=1}^{\infty} f_{ij}^{(n)}.$$

□

When $f_{ii} = 1$ then $\{f_{ii}^{(n)}\}_{n \in \mathbb{N}}$ is a probability mass function such that $f_{ii}^{(n)}$ is the probability that the M.C. first returns to state i starting from state i at time n.

Definition 4.1.7. *When $f_{ii} = 1$ then $\mu_i = \sum_{n=1}^{\infty} n f_{ii}^{(n)}$ is the* expected recurrence time *to state i.*

Definition 4.1.8.

a) State i of a Markov chain is called recurrent *if $f_{ii} = 1$ and* transient *if $f_{ii} < 1$.*

b) If state i is recurrent it is called positive recurrent *if $\mu_i < \infty$ and* null-recurrent *if $\mu_i = \infty$.*

c) A M.C. is called (positive) recurrent if all its states are (positive) recurrent.

Theorem 4.1.9. *Let $\{X_n\}$ be a homogeneous M.C. with TPM P. Then*

$$p_{ii}^{(n)} = \sum_{k=0}^{n} f_{ii}^{(k)} p_{ii}^{(n-k)} \quad \textit{for all states } i.$$

Proof. Assume the process starts in state i, so $X_0 = i$. For $1 \leq k \leq n$ let E_k be the event that $X_n = i$ and that the first return to state i is at time k. Then

$$E_j \cap E_k = \emptyset \quad \text{for} \quad j \neq k,$$

and

$$\Pr(E_k) = \Pr(\text{first return to } i \text{ at time } k) * \Pr(X_n = i \mid X_k = i) = f_{ii}^{(k)} p_{ii}^{(n-k)}$$

and so

$$p_{ii}^{(n)} = \Pr(X_n = i \mid X_0 = i) = \sum_{k=1}^{n} \Pr(E_k) = \sum_{k=1}^{n} f_{ii}^{(k)} p_{ii}^{(n-k)}$$

as claimed. □

Example 4.1.10. We can use Theorem 4.1.9 to calculate first return probabilities. Suppose a M.C. has state space $\{0, 1, 2, 3\}$ and TPM

$$P = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{pmatrix}.$$

Compute $f_{33}^{(1)}$, $f_{33}^{(2)}$, $f_{33}^{(3)}$.

Lemma 4.1.11. *Let $\{X_n\}$ be a homogeneous M.C. and let i be a statesof it. Then for any $k \in \mathbb{N}$*

$$\Pr(\#\{X_n = i\} \geq k \mid X_0 = i) = (f_{ii})^k.$$

Proof. If the process starts at state i then to hit state i at least k times it must first return to state i once and then return at least $(k-1)$ times more. Thus inductively the probability is $f_{ii}(f_{ii})^{k-1} = (f_{ii})^k$. □

In the following "i.o." means "infinitely often".

Theorem 4.1.12. *Let $\{X_n\}$ be a homogeneous M.C. and let i be a state of it. Then*

a) i transient $\Leftrightarrow \Pr(X_n = i$ i.o. $\mid X_0 = i) = 0$;

b) i recurrent $\Leftrightarrow \Pr(X_n = i$ i.o. $\mid X_0 = i) = 1$.

Proof.

$$\Pr(X_n = i \text{ i.o.} \mid X_0 = i) = \lim_{k\to\infty} \Pr(\#\{X_n = i\} \geq k \mid X_0 = k) = \lim_{k\to\infty} (f_{ii})^k$$
$$= \begin{cases} 1 & \text{if } f_{ii} = 1 \text{ (state } i \text{ recurrent)} \\ 0 & \text{if } f_{ii} < 1 \text{ (state } i \text{ transient).} \end{cases}$$

where we used Lemma 4.1.11. □

Theorem 4.1.13. *Let $\{X_n\}$ be a homogeneous M.C. and let i be a state of it. Then*

$$M_i := \sum_{n=1}^{\infty} p_{ii}^{(n)}$$

is the expected number of returns to state i. Moreover

a) i transient $\Leftrightarrow M_i < \infty$ (i.e., the expected number of returns to state i is finite).

b) i recurrent $\Leftrightarrow M_i = \infty$ (i.e., the expected number of returns to state i is infinite).

Proof. Let m_i be the random variable that counts the number of visits to state i:

$$m_i = \sum_{n=1}^{\infty} 1_{\{i\}}(X_n)$$

where

$$1_{\{i\}}(j) = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{if } j \neq i \end{cases}$$

Then the expected number of visits to state i starting from state i is

$$M_i = E(m_i \mid X_0 = i) = \sum_{n=1}^{\infty} E(1_{\{i\}}(X_n) \mid X_0 = i) = \sum_{n=1}^{\infty} p_{ii}^{(n)}.$$

Furthermore, by Lemma 4.1.11

$$\Pr(m_i \geq k \mid X_0 = i) = (f_{ii})^k, \quad k = 1, 2, 3, \ldots.$$

For any integer valued random variable Y we have

$$E(Y) = \sum_{k=1}^{\infty} k \Pr(Y = k) = \sum_{k=1}^{\infty} \Pr(Y \geq k).$$

Therefore, with $Y = m_i$,

$$M_i = \sum_{k=1}^{\infty} \Pr(m_i \geq k \mid X_0 = i) = \sum_{k=1}^{\infty} (f_{ii})^k = \begin{cases} \frac{f_{ii}}{1-f_{ii}} < \infty & \text{if } f_{ii} < 1 \quad (i \text{ transient}) \\ \infty & \text{if } f_{ii} = 1 \quad (i \text{ recurrent}) \end{cases}$$

where we used the geometric series formula in the first case. □

Example 4.1.14. Consider the random walk on $\mathbb{Z}$ with

$$\Pr(X_n = i + 1 \mid X_{n-1} = i) = p, \quad \Pr(X_n = i - 1 \mid X_{n-1} = i) = q, \quad p + q = 1.$$

Then $p_{00}^{(2n+1)} = 0$ for all $n \in \mathbb{N}_0$ since the Markov chain can only return to state 0 at even times. Moreover,

$$p_{00}^{(2n)} = \binom{2n}{n} p^n q^n = \frac{(2n)!}{n!n!} p^n q^n.$$

Stirling's approximation gives

$$n! \sim n^n e^{-n} \sqrt{2\pi n}$$

for n large (with $a_n \sim b_n$ if $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$). Using Stirling's formula we get

$$p_{00}^{(2n)} = \frac{(2n)!}{n!n!} p^n q^n \sim \frac{(2n)^{2n} \mathrm{e}^{-2n} \sqrt{4\pi n}}{n^n n^n \mathrm{e}^{-2n} \sqrt{2\pi n} \sqrt{2\pi n}} p^n q^n = \frac{(4pq)^n}{\sqrt{\pi n}}.$$

We have

$$4pq = 4p(1 - p) \leq 1$$

with equality if and only if $p = \frac{1}{2}$. To check that

$$4p(1 - p) < 1 \quad \text{if} \quad p \neq 1/2 \quad \text{and} \quad p \in [0, 1]$$

note that for

$$f(x) = 4x(1 - x)$$

we have $f(0) = f(1) = 0$ and

$$f'(x) = 4 - 8x$$

so that $f'(x) = 0$ if $x = 1/2$ and this is the maximum of f with $f(1/2) = 1$.

Case a): $p \neq q$ **(Asymmetric random walk).** Let $r = 4pq < 1$. Then for large n,

$$p_{00}^{(n)} \sim \frac{r^n}{\sqrt{n\pi}} < r^n.$$

Thus, for large m,

$$\sum_{n=m}^{\infty} p_{00}^{(n)} < \sum_{n=m}^{\infty} r^n = \frac{r^m}{1-r} < \infty.$$

Hence, by Theorem 4.1.13 the state 0 is transient for $p \neq q$.

Case b): $p = q = \frac{1}{2}$ **(Symmetric random walk).** If $p = q = 1/2$ then $4pq = 1$ and so for large m,

$$\sum_{n=m}^{\infty} p_{00}^{(n)} \sim \sum_{n=m}^{\infty} \frac{1}{\sqrt{\pi n}} = \infty$$

Hence by Theorem 4.1.13 the state 0 is recurrent. Here we used that $\sum_{n=1}^{\infty} \frac{1}{n^{\alpha}} < \infty$ if and only if $\alpha > 1$.

4.2 Irreducible Markov Chains and Communicating Classes

Definition 4.2.1. *Let i and j be states of a homogeneous M.C.. State j is called* accessible *from state i if $p_{ij}^{(n)} > 0$ for some $n \in \mathbb{N}_0$. If j accessible from i and i accessible from j then i and j are said to* communicate*, denoted by $i \leftrightarrow j$.*

Proposition 4.2.2. *$\leftrightarrow$ is an equivalence relation on the state space S of a homogeneous M.C..*

Proof. Let i, j, k be any states of the M.C.. Reflexivity $i \leftrightarrow i$ follows from the fact that $p_{ii}^{(0)} = 1$ for all i. Symmetry $i \leftrightarrow j \Leftarrow j \leftrightarrow i$ is built into the definition of $\leftrightarrow$. To prove transitivity, i.e., $i \leftrightarrow j$ and $j \leftrightarrow k$ imply $i \leftrightarrow k$ note that, since $i \leftrightarrow j$, there is some $m \in \mathbb{N}_0$ such that $p_{ij}^{(m)} > 0$, and, since $j \leftrightarrow k$, there is some $n \in \mathbb{N}_0$ such that $p_{jk}^{(n)} > 0$. This implies that

$$p_{ik}^{(m+n)} = \sum_{\ell \in S} p_{i\ell}^{(m)} p_{\ell k}^{(n)} \geq p_{ij}^{(m)} p_{jk}^{(n)} > 0, \tag{4.1}$$

where we used that $P^{(m+n)} = P^{(m)} P^{(n)}$. Therefore k is accessible from i. Exchanging the roles of i and k and arguing as before we see that i is also accessible from k and so i and k communicate. □

In words the first inequality of (4.1) can be explained as follows: one way of going from state i to state k in $(m+n)$ steps is to visit state j after m steps and then state k after another n steps. The probability of this event is $p_{ij}^{(m)} p_{jk}^{(n)}$.

Definition 4.2.3. *We can partition the state space into disjoint equivalence classes which we call* communicating classes.

Remark 4.2.4. *A Markov chain might start in a communicating class C_1 and then enter another communicating class C_2, but then it cannot return to C_1.*

Definition 4.2.5. *A M.C. is called* irreducible *if it only has one communicating class.*

Example 4.2.6. The Markov chain with TPM

$$P = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$$

on $S = \{1, 2, 3\}$ is irreducible.

Example 4.2.7. If a Markov chain with state space S has TPM

$$P = \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix},$$

i.e., P, is block-diagonal such that $S = C_1 \cup C_2$ with $P_1 = P|_{C_1}$ and $P_2 = P|_{C_2}$ and C_1, C_2 are communicating classes then the M.C. stays in C_1 forever if it starts in C_1 and stays in C_2 forever if it starts in C_2.

Definition 4.2.8. *A communicating class C is called* closed *if*

$$\Pr(X_n \in C \mid X_0 \in C) = 1.$$

Example 4.2.9. Determine the closed communicating classes of the random walk on the integers $\{0, 1, \dots, N\}$ with absorbing boundaries at site 0 and N.

Remark 4.2.10. *A closed communicating class of a M.C. with TPM P on the state space S is a M.C. itself. To see this let $C \subseteq S$ be a communicating class of the Markov chain. Then $\sum_{k \in C} p_{ik} = 1$ for all $i \in C$ since $p_{ir} = 0$ for all $r \notin C$. Thus C has its own TPM given by $P|_C$.*

Theorem 4.2.11. (Invariants under communication) *Let i and j be states of a homogeneous M.C. with TPM P. Then the following holds:*

a) If $i \leftrightarrow j$ then i and j have the same period.

b) If $i \leftrightarrow j$ then i is recurrent if and only if j is recurrent.

Proof.

a) Let $i \leftrightarrow j$. Recall that the period $\tau(i)$ of state i is the greatest number which divides all its possible return times, i.e., $\tau(i) = \gcd(n \geq 1, p_{ii}^{(n)} > 0)$. Since $i \leftrightarrow j$ there are $r, s \in \mathbb{N}_0$ such that $p_{ij}^{(r)} > 0$ such that $p_{ji}^{(s)} > 0$. Hence by (4.1)

$$p_{jj}^{(r+s)} \geq p_{ji}^{(s)} p_{ij}^{(r)} > 0,$$

and so $r + s$ is a possible return time for state j and hence $\tau(j)$ divides $r + s$. Next note that for $r, s, n \in \mathbb{N}_0$ we have

$$p_{jj}^{(r+n+s)} \geq p_{ji}^{(s)} p_{ii}^{(n)} p_{ij}^{(r)}, \tag{4.2}$$

because (similarly to the proof of (4.1)) one way of returning to state j after $r + n + s$ steps is to visit state j after s steps (the probability of this is $p_{ji}^{(s)}$), then return to state i after n steps (the probability of this is $p_{ii}^{(n)}$) and then to visit state j after r steps (the probability of this is $p_{ij}^{(r)}$), with the probability of the whole event given as the product $p_{ji}^{(s)} p_{ii}^{(n)} p_{ij}^{(r)}$ of the probabilities.

Now let $n \in \mathbb{N}$ be a possible return time for state i, so that $p_{ii}^{(n)} > 0$; then by (4.2)

$$p_{jj}^{(r+n+s)} \geq p_{ji}^{(s)} p_{ii}^{(n)} p_{ij}^{(r)} > 0,$$

and so $n+r+s$ is a possible return time for state j as well; therefore $\tau(j)$ divides $n+r+s$. Since $\tau(j)$ also divides $r+s$ it divides n. Hence $\tau(j)$ divides all possible return times n of state i. But $\tau(i)$ is the largest integer with the property that it divides all possible return times $n \in \mathbb{N}$ of state i. Hence $\tau(i) \geq \tau(j)$. Exchanging the role of i and j shows that $\tau(j) \geq \tau(i)$ and so $\tau(i) = \tau(j)$.

b) Suppose state i is recurrent. By Theorem 4.1.13 this is equivalent to

$$M_i = \sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty,$$

where M_i is the expected number of visits to state i. Suppose $i \leftrightarrow j$. Then we need to prove that

$$M_j = \sum_{n=1}^{\infty} p_{jj}^{(n)} = \infty.$$

Since $i \leftrightarrow j$ there are $r \in \mathbb{N}$ such that $p_{ij}^{(r)} > 0$ and $s \in \mathbb{N}$ such that $p_{ji}^{(s)} > 0$ (as in part a)). Hence (4.2) gives

$$p_{jj}^{(s+n+r)} \geq p_{ji}^{(s)} p_{ii}^{(n)} p_{ij}^{(r)} > 0,$$

and so

$$\sum_{\ell=1}^{\infty} p_{jj}^{(\ell)} \geq \sum_{\ell=r+s+1}^{\infty} p_{jj}^{(\ell)} \geq \sum_{n=1}^{\infty} p_{ji}^{(s)} p_{ii}^{(n)} p_{ij}^{(r)} = p_{ji}^{(s)} \left(\sum_{n=1}^{\infty} p_{ii}^{(n)} \right) p_{ij}^{(r)} = p_{ji}^{(s)} M_i p_{ij}^{(r)} = \infty.$$

Hence by Theorem 4.1.13 state j is recurrent. So we have proved that i recurrent $\Rightarrow$ j recurrent. Exchanging the roles of i and j in the above argument also shows j recurrent $\Rightarrow$ i recurrent.

□

Theorem 4.2.12. *Any finite homogeneous irreducible M.C. $\{X_n\}$ is recurrent.*

Proof. Since the M.C. is irreducible either all states are recurrent or all states are transient by Theorem 4.2.11. Suppose by contradiction that all states are transient. Let $X_0 = i_0$. By transience $\Pr(X_n = i_0 \text{ i.o.}) = 0$ due to Theorem 4.1.12. So there is $n_0 \in \mathbb{N}$ such that at time n_0 is the last visit to state i_0. Let $X_{n_0+1} = i_1$. Then $i_1 \neq i_0$. By transience, $\Pr(X_n = i_1 \text{ i.o.}) = 0$. So there is $n_1 \geq n_0 + 1 \in \mathbb{N}$ such that at time n_1 is the last visit to state i_1. Then $X_n \neq i_0, i_1$ for $n > n_1$. Let $X_{n_1+1} = i_2$. Then $i_2 \neq i_0, i_1$. Continuing like this, after finitely many steps we have excluded all possibilities which is a contradiction. □

Corollary 4.2.13. *Finite closed communicating classes are Markov chains themselves (see Remark 4.2.10) and hence recurrent.*

Example 4.2.14. Consider a Markov chain on the state space $S = \{0, 1, 2, 3, 4, 5\}$ with TPM

$$P = \begin{pmatrix} \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \end{pmatrix}$$

Find the recurrent communicating classes and compute the periods of all states.

Proposition 4.2.15. *Recurrent communicating classes of homogeneous Markov chains are closed.*

Proof. We prove the contraposition: non-closed communicating classes are transient. Let C be a non-closed communicating class. . Then there are $j \notin C$, $i \in C$ such that j is accessible from i. If C was a recurrent communicating class then i would be recurrent, gebce state i would have to be be accessible from state j. But this would imply $i \leftrightarrow j$ and so $j \in C$ which is a contradiction. Thus C is transient. □

Theorem 4.2.16. *A M.C. can be partitioned into disjoint classes $T, C_1, C_2, \ldots$ such that T consists of transient states and the C_i are closed recurrent communicating classes.*

Proof. Any state i in the Markov chain is either transient or recurrent. If it is recurrent then it belongs to a recurrent communicating class (by Theorem 4.2.11 b)) and by Proposition 4.2.15 this communicating class is closed. □

Definition 4.2.17. *A state j of a homogeneous Markov chain is called* ergodic *if it is recurrent and aperiodic.*

Proposition 4.2.18. *Let i, j be two states of a homogeneous Markov chain. If $i \leftrightarrow j$ then i ergodic $\Leftrightarrow$ j ergodic.*

Proof. follows from Theorem 4.2.11 (invariants under communication). □

Definition 4.2.19. *A communicating class C of a homogeneous Markov chain is called ergodic if all its states are ergodic.*

Theorem 4.2.20. *Let $\{X_n\}$ be a finite, irreducible, homogeneous M.C.. Then*

$$\{X_n\} \quad \textit{is ergodic} \quad \Leftrightarrow \quad \{X_n\} \quad \textit{is aperiodic}.$$

Proof. The recurrence follows from Theorem 4.2.12. □

Definition 4.2.21. *Let $\{X_n\}$ be a homogeneous Markov chain. Then $T_j = \inf(n \in \mathbb{N}, X_n = j)$ is called* first passage time *of state j.*

If $X_0 = j$ we also call T_j first return time of state j. Note that

$$f_{jj} = \Pr(T_j < \infty \mid X_0 = j)$$

and so $\Pr(T_j < \infty \mid X_0 = j) = 1$ if and only if j is a recurrent state.

The next theorem shows that in a recurrent irreducible M.C. every state is visited by the process with probability one.

Theorem 4.2.22. *Let $\{X_n\}$ be a homogeneous M.C. with state space S.*

a) *Let $i, j \in S$ and assume that state j is recurrent and $i \leftrightarrow j$. Then*

$$f_{ij} := \Pr(T_j < \infty \mid X_0 = i) = 1.$$

b) *Assume that the M.C. is irreducible and recurrent. Then for all states $j \in S$ we have $T_j < \infty$.*

Proof in words:

a) Since i is accessible from j it is possible that the Markov chain visits j starting from i ($p_{ji}^{(n)} > 0$ for some $n \in \mathbb{N}$). But the Markov chain visits state j infinitely often by Theorem 4.1.12 since state j is recurrent. Therefore the Markov chain has to returrn to state j starting from i. This means that $f_{ij} = 1$.

b) By a), since the Markov chain is now assumed irreducible and recurrent, whatever state $i \in S$ the Markov chain is in at time 0 it will visit every state j at some time T_j. Hence $T_j < \infty$ for all $j \in S$.

□

Formal Proof.[1]

a) Since $i \leftrightarrow j$ there is some m such that $p_{ji}^{(m)} > 0$. Since state j is recurrent we have

$$\begin{aligned}
1 &= \Pr(X_n = j \quad \text{i.o.} \mid X_0 = j) \\
&\leq \Pr(X_n = j \quad \text{for some } \ n \geq m+1 \mid X_0 = j) \\
&= \sum_{k \in S} \Pr(X_n = j \quad \text{for some } \ n \geq m+1 \mid X_m = k, X_0 = j) \Pr(X_m = k \mid X_0 = j) \\
&= \sum_{k \in S} f_{kj} p_{jk}^{(m)}.
\end{aligned}$$

So we have both

$$\sum_{k \in S} f_{kj} p_{jk}^{(m)} = 1, \quad \sum_{k \in S} p_{jk}^{(m)} = 1.$$

For the first equation to hold true we therefore need $f_{kj} = 1$ whenever $p_{jk}^{(m)} \neq 0$, in particular this holds for $k = i$.

b) We have $T_j < \infty$ if $\Pr(T_j < \infty) = 1$ and

$$\begin{aligned}
\Pr(T_j < \infty) &= \sum_{i \in S} \Pr(T_j < \infty \mid X_0 = i) \Pr(X_0 = i) = \sum_{i \in S} f_{ij} \Pr(X_0 = i) \\
&= \sum_{i \in S} \Pr(X_0 = i) = 1,
\end{aligned}$$

where we have used part a).

□

[1]Material not covered in lectures and not examinable

4.3 Stationary Distributions of Markov Chains

Definition 4.3.1. *Given a homogeneous M.C.* $\{X_n\}$ *on a state space* $S = \{0, 1, \ldots, N\}$, $N \in \mathbb{N} \cup \{\infty\}$, *with TPM* P *a probabiity mass function* $\pi = \{\pi_j\}_{j \in S}$ *is called* stationary distribution *for the Markov chain if if* $\pi P = \pi$, *which is in coordinates*

$$\sum_{j=0}^{N} \pi_i p_{ij} = \pi_j, \quad j = 0, 1, \ldots, N.$$

Remarks 4.3.2.

a) For a stationary initial distribution $\Pr(X_0 = i) = \pi_i$, $i \in S$ *we have*

$$\Pr(X_n = i) = \pi_i \quad \textit{for all } n \in \mathbb{N},\ i \in S,$$

because by Corollary 3.3.6 for all $i \in S$

$$\Pr(X_n = i) = (\pi P^n)_i$$

and by the definition of a stationary distribution

$$(\pi P^n)_i = \pi_i, \quad \textit{for all } \ i \in S.$$

In other words, a stationary distribution π *does not change with time.*

b) The column vector π^T *corresponding to the row vector of the stationary distribution* $\pi = (\pi_0, \pi_1, \ldots, \pi_N)$ *of a homogeneous M.C. on a state space* $S = \{0, 1, \ldots, N\}$ *with TPM* P *satisfies*

$$\pi^T = P^T \pi^T,$$

so that π *lies in the null space of* $P^T - \mathrm{id}$ *(where* id *is the identity matrix). To find* π *we solve* $\pi^T = P^T \pi^T$ *together with the constraint*

$$\pi_0 + \pi_1 + \ldots + \pi_N = 1.$$

Example 4.3.3. Compute the stationary distribution of the Markov chain with state space $S = \{1, 2, 3\}$ and TPM

$$P = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Example 4.3.4. Clearly the Markov chain with TPM

$$P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

has infinitely many stationary distributions.

In this section we will find conditions which guarantee a stationary distribution to exist and to be unique.

Definition 4.3.5. *Let $\{X_n\}$ be a homogeneous Markov chain with finite or infinite state space $S = \{0, 1, 2, \dots, N\}$ where $N \in \mathbb{N} \cup \{\infty\}$. If there exists a probability distribution π on S with*

$$\lim_{n\to\infty} \Pr(X_n = i) = \pi_i$$

for all initial distributions $p = \{p_j = \Pr(X_0 = j)\}_{j\in S}$ then π is called the limiting distribution *of the Markov chain.*

Proposition 4.3.6. *If a homogeneous Markov chain $\{X_n\}$ has a limiting distribution π then π is a stationary distribution.*

Proof. A limiting distribution π satisfies $\pi_i = \lim_{n\to\infty} \Pr(X_n = i)$ for all $i \in S$ and so does not change after a time step: we have $\lim_{n\to\infty} \Pr(X_{n+1} = i) = \pi_i$ as well. Formally we compute, using Corollary 3.3.6, that

$$\pi_j = \lim_{n\to\infty} \Pr(X_n = j) = \lim_{n\to\infty} (pP^n)_j = \lim_{n\to\infty} (pP^{n+1})_j = ((\lim_{n\to\infty} pP^n)P)_j = (\pi P)_j$$

for all $j \in S$ and so $\pi P = \pi$. □

Definition 4.3.7. *Let $\{X_n\}$ be a homogeneous, irreducible, recurrent M.C. and let i, j be states of it. Let, as before, $1_{\{i\}}(k) = 1$ if $k = i$ and 0 otherwise. Define*

$$\gamma_i^j = E\Big(\sum_{n=0}^{T_j-1} 1_{\{i\}}(X_n) \mid X_0 = j\Big),$$

where T_j is the first return time to state j, see Definition 4.2.21.

Note that γ_i^j is the expected number of visits to state i between visits to state j for $i \neq j$. Moreover for $i = j$

$$\gamma_j^j = E\Big(\sum_{n=0}^{T_j-1} 1_{\{j\}}(X_n) \mid X_0 = j\Big) = 1_{\{j\}}(j) + E\Big(\sum_{n=1}^{T_j-1} 1_{\{j\}}(X_n) \mid X_0 = j\Big) = 1 + 0 = 1.$$

Theorem 4.3.8. *Let $\{X_n\}$ be a homogeneous, irreducible, recurrent M.C. with state space $S = \{0, 1, \dots, N\}$, $N \in \mathbb{N} \cup \{\infty\}$, and let $j \in S$. Let $\gamma^j = (\gamma_0^j, \gamma_1^j, \dots, \gamma_N^j)$. Then*

a) $\gamma^j P = \gamma^j$;

b) $0 < \gamma_i^j < \infty$ for all $i, j \in S$.

Proof.[1]

a) We will not prove part a) here, for a proof see the book "Markov Chains" by Norris. But note that it is reasonable that γ_i^j, the expected number of visits to a state i in between visits to j is invariant under time evolution which gives $\gamma^j P = \gamma^j$.

b) To prove b) notice that by Theorem 4.2.22 we have $T_j < \infty$ and therefore, due to the definition of γ_i^j also $\gamma_i^j < \infty$ for all i, j. We will not prove here that $\gamma_i^j > 0$ (a proof can be found in the book "Markov Chains" by Norris); but note that we already know that $\gamma_j^j = 1$. Moreover any state $i \neq j$ is accessible from state j so that it is reasonable that the expected number γ_i^j of visits to state j between visits to state i is not zero.

[1] Material not covered in lectures and not examinable

□

Theorem 4.3.9. *An irreducible, recurrent homogeneous M.C. $\{X_n\}$ on the state space S with a positive recurrent state j has a stationary distribution $\pi = \frac{1}{\mu_j}\gamma^j$ with $\pi_i > 0$ for all $i \in S$. In particular $\pi_j = \frac{1}{\mu_j}$.*

Proof. Let $S = \{0, 1, \ldots, N\}$ be the state space of the Markov chain. Then

$$\begin{aligned}\sum_{i=0}^{N} \gamma_i^j &= \text{expected number of visits to other states between visits to } j \\ &= \text{expected time between visits to } j = \text{expected return time to } j = \mu_j\end{aligned}$$

and

$$\mu_j < \infty \text{ if and only if } j \text{ is positively recurrent.}$$

Since $\mu_j > 0$ by definition we can define $\pi = \frac{1}{\mu_j}\gamma^j$. Since $\mu_j < \infty$ we then know that π is not the zero vector, i.e., $\pi \neq 0$. Moreover $\pi P = \pi$ by Theorem 4.3.8. Since by definition $\gamma_i^j \geq 0$ we have $\pi_i = \gamma_i^j/\mu_j \geq 0$ for all $i \in S$. Moreover

$$\sum_{i=0}^{N} \pi_i = \frac{1}{\mu_j} \sum_{i=0}^{N} \gamma_i^j = 1.$$

Further, since $\gamma_j^j = 1$ we have $\pi_j = \gamma_j^j/\mu_j = 1/\mu_j$. Note that indeed $\pi_i = \gamma_i^j/\mu_j > 0$ for all $i \in S$ because $\gamma_i^j > 0$ for all $i \in S$ by Theorem 4.3.8 and because by assumption $\mu_j < \infty$. □

Theorem 4.3.10. *A finite irreducible homogeneous M.C. has a stationary distribution.*

Proof. Let $S = \{0, 1, \ldots, N\}$, where $N \in \mathbb{N}_0$, be the state space of the Markov chain. By Theorem 4.2.12 all states are recurrent. Moreover by Theorem 4.3.8 we know that $0 < \gamma_i^j < \infty$ for all $i, j \in S$. Since N is finite and $\mu_j = \sum_{i=0}^{N} \gamma_i^j$ we see that $\mu_j < \infty$ and so state j is positive recurrent. Therefore by Theorem 4.3.9 the Markov chain has a stationary distribution. □

Theorem 4.3.11. *An irreducible, recurrent, homogeneous M.C. has at most one stationary distribution π and if it exists then $\pi_j = \frac{1}{\mu_j} > 0$ for all j.*

Proof.[1] The proof is in the book "Markov chains" by Norris.

Corollary 4.3.12. *If an irreducible, homogeneous M.C. has a positive recurrent state j, then all states are positive recurrent.*

Proof By Theorem 4.3.9 the Markov chain has a stationary distribution π and by Theorem 4.3.11 is is given by $\pi_i = \frac{1}{\mu_i}$ with $\pi_i > 0$ for all $i \in S$. Since $\pi_i > 0$ if and only if $\mu_i < \infty$ state i is positive recurrent for all $i \in S$. □

Corollary 4.3.13. *A positive recurrent, irreducible homogeneous M.C. has exactly one stationary distribution π with $\pi_j = \frac{1}{\mu_j} > 0$ for all states j.*

Proof By Theorem 4.3.9 the Markov chain has a stationary distribution π and by Theorem 4.3.11 it is unique and given by $\pi_i = \frac{1}{\mu_i}$ for all $i \in S$. □

Examples 4.3.14.

a) Any finite irreducible homogeneous Markov chain has a unique stationary distribution (by Theorems 4.3.10 and 4.3.11)

b) The symmetric random walk is irreducible and recurrent (see Example 4.1.14). Therefore by Theorem 4.3.8 there is a sequence γ^j with components γ_i^j satisfying $0 < \gamma_i^j < \infty$ and $\gamma^j P = \gamma^j$. Suppose that there is also a stationary distribution π. Then $\pi P = \pi$ is equivalent to $\pi_i = \sum_{j=-\infty}^{\infty} \pi_j p_{ji}$. Since $p_{i,i+1} = \frac{1}{2} = p_{i,i-1}$ for the symmetric random walk and $p_{ij} = 0$ for $j \neq i \pm 1$, this gives
$$\pi_j = \frac{1}{2}\pi_{j-1} + \frac{1}{2}\pi_{j+1}.$$
Moreover, $\sum_{k=-\infty}^{\infty} \pi_k = 1$. To solve this difference equation try $\pi_j = \lambda^j$. Then, dividing by λ^{j-1}, assuming $\lambda \neq 0$ we get $\lambda = \frac{1}{2}(1 + \lambda^2)$ and $\lambda = 1$ is a double root of this equation. So the general solution of this difference equation is $\pi_k = c_1 + c_2 k$. Since $0 \leq \pi_k \leq 1$ for all k we have $c_2 = 0$. But then π is given by $\pi_i \equiv c_1$ and so does not satisfy $\sum_{k=-\infty}^{\infty} \pi_k = 1$. Therefore the symmetric random walk does not have a stationary distribution and is therefore null-recurrent.

4.4 The Basic Limit Theorem for Markov Chains

Theorem 4.4.1. *(Basic Limit Theorem, [BLT]) An irreducible, positive recurrent, aperiodic homogeneous M.C. with TPM P and state space $\{0, 1, \ldots, N\}$, $N \in \mathbb{N} \cup \{\infty\}$, has a distribution π such that*
$$\lim_{n\to\infty} p_{ji}^{(n)} = \pi_i \quad \textit{for all} \quad j, i.$$

A proof can be found in the book "Markov chains" by Norris.

Remarks 4.4.2.

a) The distribution π from the BLT is a limiting distribution, i.e., for any initial distribution p (i.e., $\Pr(X_0 = i) = p_i$, $i = 0, \ldots, N$) we have
$$\lim_{n\to\infty} \Pr(X_n = i) = \pi_i.$$
This follows from
$$\lim_{n\to\infty} \Pr(X_n = i) = \lim_{n\to\infty} (pP^n)_i = \lim_{n\to\infty} \sum_{j=0}^{N} p_j p_{ji}^{(n)} = \sum_{j=0}^{N} p_j \pi_i = \pi_i,$$
where we used Corollary 3.3.6 in the first equation. □

b) By Proposition 4.3.6 the limiting distribution is stationary, and by Theorem 4.3.11 it is given by $\pi_j = \frac{1}{\mu_j}$, $j = 0, \ldots, N$.

Theorem 4.4.3. *Let P be the TPM of a homogeneous M.C. $\{X_n\}$ and assume that*
$$\lim_{n\to\infty} p_{ji}^{(n)} = \frac{1}{\mu_i}$$
for all states i. Then the M.C. spends an average fraction $\frac{1}{\mu_i}$ in state i.

Proof. Let $m_i^{(n)}$ be the random variable that counts the number of visits of the M.C. to state i during the times $0, 1, \ldots, n-1$. Then

$$m_i^{(n)} = \sum_{k=0}^{n-1} 1_{\{i\}}(X_k)$$

where, as before,

$$1_{\{i\}}(j) = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{otherwise.} \end{cases}$$

Then, as we saw before (with $i = j$, see proof of Theorem 4.1.13)

$$E(m_i^{(n)} \mid X_0 = j) = \sum_{k=0}^{n-1} E(1_{\{i\}}(X_k) \mid X_0 = j) = \sum_{k=0}^{n-1} p_{ji}^{(k)}.$$

Thus,

$$E(\frac{m_i^{(n)}}{n} \mid X_0 = j) = \frac{1}{n}\sum_{k=0}^{n-1} p_{ji}^{(k)}$$

is the expected fraction of time spent in state i in the period of time $0, 1, \ldots, n-1$ if the process starts in state j. Since $\lim_{n\to\infty} p_{ji}^{(n)} = \frac{1}{\mu_i}$ we conclude that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} p_{ji}^{(k)} = \frac{1}{\mu_i},$$

where we used that if $\lim_{n\to\infty} a_n = a$ for a sequence $\{a_n\}_{n\in\mathbb{N}}$ then

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} a_n = a.$$

Thus, in the long run, the M.C. spends a fraction of time $\frac{1}{\mu_i}$ in state i. □

Example 4.4.4. Suppose an irreducible, aperiodic, recurrent, homogeneous M.C. on $S = \{0, 1, 2\}$ satisfies

$$\mu_0 = \mu_1 = 4, \ \mu_2 = 2.$$

Then the BLT applies and so, by Theorem 4.4.3, on average the M.C. spends half of its time in state 2, a quarter in state 0 and a quarter in state 1.

Example 4.4.5. A Markov chain with state space $\{1, 2, 3\}$ has transition probability matrix

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & 1/2 \end{pmatrix}.$$

Compute the stationary distribution of this Markov chain. The costs incurred in spending one unit period are £2 in state 1, £1 in state 2 and £3 in state 3. What is the long run cost per unit period of this Markov chain?

Remark 4.4.6. *A finite closed communicating class C of a homogeneous M.C. is a M.C. itself (see Remark 4.2.10) and so by Corollary 4.3.13 it has a unique stationary distribution π with $\pi_i = 1/\mu_i > 0$ for all $i \in C$. If C is aperiodic, then the BLT (Theorem 4.4.1) and Theorem 4.4.3 hold too.*

Chapter 5

Poisson Processes

The Poisson distribution and Poisson processes come up frequently in applications when studying "rare events", e.g. break downs in transmissions, traffic accidents, arrivals at queues etc.

5.1 Poisson Distribution

Definition 5.1.1. *The* Poisson distribution *with parameter* $\mu \geq 0$ *is the probability mass function given by* $p_k = e^{-\mu}\mu^k/k!$, $k \in \mathbb{N}_0$. *A random variable* X *with values in* $\mathbb{N}_0$ *is Poisson distributed, if* $\Pr(X = k) = e^{-\mu}\mu^k/k!$.

Proposition 5.1.2. *The Poisson distribution is indeed a probability mass function, i.e.,*

a) $0 \leq p_k \leq 1$ *for all* $k \in \mathbb{N}_0$*;*

b) $\sum_{k=0}^{\infty} p_k = 1$.

Proof. We use the Taylor expansion of e^{μ}:

$$e^{\mu} = 1 + \mu + \frac{\mu^2}{2!} + \frac{\mu^3}{3!} + \dots$$

a) It is clear that $p_k \geq 0$. To prove that $p_k \leq 1$ note that

$$p_k = \frac{\mu^k}{k!}e^{-\mu} \leq (1 + \mu + \frac{\mu^2}{2!} + \frac{\mu^3}{3!} + \dots)e^{-\mu} = e^{\mu} \cdot e^{-\mu} = 1.$$

b) $\sum_{k=0}^{\infty} p_k = \sum_{k=0}^{\infty} \frac{\mu^k}{k!} e^{-\mu} = e^{\mu} \cdot e^{-\mu} = 1.$

□

Proposition 5.1.3. *Let* X *be a random variable which is Poisson distributed with parameter* μ*. Then* $E(X) = \mu$*.*

Proof.

$$E(X) = \sum_{k=0}^{\infty} kp_k = \sum_{k=0}^{\infty} k\frac{\mu^k}{k!}e^{-\mu} = \mu e^{-\mu}\sum_{k=1}^{\infty}\frac{\mu^{k-1}}{(k-1)!} = \mu e^{-\mu}\sum_{k=0}^{\infty}\frac{\mu^k}{k!} = \mu e^{-\mu}e^{\mu} = \mu.$$

□

5.2 Law of Rare Events

When an event can occur in a large number of independent ways and the probability of each such event is the same arbitrarily small number then the number of events follows a Poisson distribution:

Theorem 5.2.1. *Consider n independent trials, probability of success in each trial is p; let $X_{n,p}$ be the number of successes in n trials. Keep $\mu = np = E(X_{n,p})$ constant and $n \to \infty$. Then*

$$\lim_{\substack{n\to\infty \\ \mu=np}} \Pr(X_{n,p} = k) = \frac{\mu^k e^{-\mu}}{k!}.$$

Proof. We have

$$\begin{aligned}
\Pr(X_{n,p} = k) &= \binom{n}{k} p^k (1-p)^{n-k} \\
&= n(n-1)\cdot\ldots(n-k+1)\frac{p^k(1-p)^n}{k!(1-p)^k} \\
&= n(n-1)\cdot\ldots(n-k+1)\frac{(\frac{\mu}{n})^k(1-\frac{\mu}{n})^n}{k!(1-\frac{\mu}{n})^k} \\
&= 1(1-\frac{1}{n})\cdot\ldots\cdot(1-\frac{k-1}{n})\frac{\mu^k(1-\frac{\mu}{n})^n}{k!(1-\frac{\mu}{n})^k}
\end{aligned}$$

and so

$$\lim_{n\to\infty} \Pr(X_{n,p=\frac{\mu}{n}} = k) = \frac{\mu^k}{k!}\lim_{n\to\infty}(1-\frac{\mu}{n})^n = \frac{\mu^k e^{-\mu}}{k!}$$

which is the Poisson distribution. □

5.3 Poisson Processes

A Poisson process $X(t)$ models that number of rare events up to time t and is characterized by the rate of occurence of rare events $\lambda \geq 0$. It is used to model for example disintegration of radioactive particles, occurence of accidents etc.

Definition 5.3.1. *A* Poisson process *of rate $\lambda > 0$ is an integer valued stochastic process $\{X_t\}_{t\geq 0}$ for which*

(i) for any times $t_0 < t_1 < \ldots < t_n < \ldots$ the process increments

$$X(t_1) - X(t_0), X(t_2) - X(t_1), \ldots, X(t_n) - X(t_{n-1}), \ldots$$

are independent.

(ii) For $s \geq 0$, $t > 0$ the random variable $X(s+t) - X(s)$ is Poisson distributed with parameter λt,

$$\Pr(X(s+t) - X(s) = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}, \quad k = 0, 1, 2, \ldots$$

(iii) $X(0) = 0$.

Remarks 5.3.2.

a) *From* $X(0) = 0$ *and (ii) we get*

$$\Pr(X(t) = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!},$$

b) *Note also that* $E(X(t)) = \lambda t$ *by Proposition 5.1.3 with* $\mu = \lambda t$.

Example 5.3.3. Defects occur along a communication cable according to a Poisson process at rate $\lambda = 0.1$ per minute.

a) What is the probability that no defects occur in the first two minutes of operation?

b) Suppose it is known that there we no defects in the first two minutes. What is the probability of no defects between $t = 2$ and $t = 3$?

5.4 Derivation of the Poisson Process

We have constructed the Poisson process from the Poisson distribution. We can also derive it from the postulates below using First step analysis.

Definition 5.4.1.

a) $f : \mathbb{R} \to \mathbb{R}$ *is* $o(h)$ *if* $\lim_{h\to 0} f(h)/h = 0$.

b) $f(h)$ *is* $O(h)$ *if there is* $c \in \mathbb{R}$ *with* $|f(h)/h| \leq c$ *as* $h \to 0$.

Note that $o(h) - o(h) = o(h)$, it is in general not 0! Also $-o(h) = o(h)$.

Example 5.4.2.

a) Let $f(h) = h^{3/2}$. Is $f(h)$ $O(h)$ or $o(h)$ or none?

b) Let $f(h) = 3h$. Is $f(h)$ $O(h)$ or $o(h)$ or none?

Postulates 5.4.3. *(Poisson process)*
Let $N((a, b])$ *be the number of events occuring during the time interval* $(a, b]$.

1) *For* $t_0 = 0 < t_1 < \ldots < t_m$ *the random variables* $N((t_0, t_1]), N((t_1, t_2]), \ldots, N((t_{m-1}, t_m])$ *are independent.*

2) *For any* $t > 0, h > 0$, *the probability distribution of* $N((t, t+h])$ *(# events occuring between time* t *and* $t + h$*) depends only on the interval length* h, *not on* t.

3) $\exists \lambda > 0$ *such that* $\Pr(N((t, t + h]) \geq 1) = \lambda h + o(h)$ *as* $h \to 0$.

4) $\Pr(N((t, t + h]) \geq 2) = o(h)$ *as* $h \to 0$.

Theorem 5.4.4. *Let* $P_n(t) = \Pr(N((0, t]) = n)$. *The postulates above imply that*

$$P_n(t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}$$

i.e., $P_n(t)$ *is Poisson distributed with parameter* $\mu = \lambda t$, *and* $X(t) = N((0, t])$ *is a Poisson process.*

Proof. First step analysis gives for $n \geq 1$

$$P_n(t+h) = P_n(t)P_0(h) + P_{n-1}(t)P_1(h) + \sum_{i=2}^{n} P_{n-i}(t)P_i(h). \tag{5.1}$$

By postulate 3

$$P_0(h) = 1 - \Pr(N((t,t+h]) \geq 1) = 1 - \lambda h + o(h)$$

and

$$\Pr(N((t,t+h]) \geq 1) = \lambda h + o(h) = \Pr(N((t,t+h]) = 1) + \Pr(N((t,t+h]) \geq 2) = P_1(h) + o(h)$$

where we used postulate 4. Hence

$$P_1(h) = \lambda h + o(h).$$

Finally we can estimate the third term of (5.1) as follows:

$$\sum_{i=2}^{n} P_{n-i}(t)P_i(h) \leq \sum_{i=2}^{n} P_i(h) \leq \sum_{i=2}^{\infty} P_i(h) = \Pr(N((t,t+h]) \geq 2) = o(h),$$

where we have used postulate 4. Substituting this into (5.1) gives

$$P_n(t+h) = (1-\lambda h)P_n(t) + P_{n-1}(t)\lambda h + o(h)$$

and

$$\frac{P_n(t+h) - P_n(t)}{h} = -\lambda P_n(t) + P_{n-1}(t)\lambda + \frac{o(h)}{h}$$

for $n \geq 1$. Letting $h \to 0$ we get

$$\frac{\mathrm{d}P_n(t)}{\mathrm{d}t} = -\lambda P_n(t) + \lambda P_{n-1}(t), \quad n \geq 1,$$

If $n = 0$ then the second term in this ODE has to be dropped (since the number of events is always non-negative). So we get

$$P_0(t+h) = P_0(t)(1 - \lambda_0 h + o(h))$$

and

$$\frac{P_0(t+h) - P_0(t)}{h} = -\lambda P_0(t) + \frac{o(h)}{h}$$

Letting $h \to 0$ we get

$$\frac{\mathrm{d}P_0(t)}{\mathrm{d}t} = -\lambda P_0(t)$$

This gives the system of ODEs

$$\frac{\mathrm{d}P_n(t)}{\mathrm{d}t} = -\lambda P_n(t) + \lambda P_{n-1}(t), \quad n \geq 1,$$
$$\frac{\mathrm{d}P_0(t)}{\mathrm{d}t} = -\lambda P_0(t)$$

with initial values

$$P_0(0) = 1, \quad P_n(0) = 0 \text{ for } n \geq 1$$

(this follows from $P_n(0) = \Pr(N((0,0]) = n) = \Pr(N(\emptyset) = n)$ which is 0 as no events happen in an empty time interval. Hence $P_0(t) = e^{-\lambda t}$ and

$$P_1'(t) = -\lambda P_1(t) + \lambda e^{-\lambda t}.$$

Multiplying through with $e^{\lambda t}$ we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}(e^{\lambda t} P_1(t)) = \lambda$$

and so

$$e^{\lambda t} P_1(t) = \lambda t + c$$

and

$$P_1(t) = \lambda t e^{-\lambda t} + c e^{-\lambda t}$$

Since $P_1(0) = 0$ we have $P_1(t) = \lambda t e^{-\lambda t}$. This proves the theorem for $n = 1$. Now assume that the theorem holds for some $n \in \mathbb{N}$, i.e., $P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$. Then

$$\frac{\mathrm{d}P_{n+1}(t)}{\mathrm{d}t} = -\lambda P_{n+1}(t) + \lambda P_n(t) = -\lambda P_{n+1}(t) + \lambda \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

which is equivalent to

$$\frac{\mathrm{d}(e^{\lambda t} P_{n+1}(t))}{\mathrm{d}t} = \frac{\lambda^{n+1} t^n}{n!} \quad \Longrightarrow \quad e^{\lambda t} P_{n+1}(t) = \frac{\lambda^{n+1} t^{n+1}}{(n+1)!} + c$$

Since $P_{n+1}(0) = 0$ we have $c = 0$ and so $P_{n+1}(t) = \frac{\lambda^{n+1} t^{n+1}}{(n+1)!} e^{-\lambda t}$ as claimed.

□

Remark 5.4.5. *The postulates for the Poisson process imply*

1) $\Pr(X(t+h) - X(t) = 1 \mid X(t) = j) = \lambda h + o(h)$ *for all* $j \in \mathbb{N}_0$*;*

2) $\Pr(X(t+h) - X(t) = 0 \mid X(t) = j) = 1 - \lambda h + o(h)$ *for all* $j \in \mathbb{N}_0$*;*

3) $X(0) = 0$.

5.5 Examples of Poisson Processes

Example 5.5.1. (Radioactive decay) *Let* $X(t)$ *be the number of radioactive disintegrations detected by a counter in the time interval* $[0, t]$*. The process is Poisson as long as the half life time of the substance is large compared to* $[0, t]$*.*

Example 5.5.2. (Fishing) *Let* $X(t)$ *be the number of fish caught in the time interval* $[0, t]$*. Assume that the number of fish is large and that fish are equally likely to bite at any time. Then* $\{X(t)\}$ *may be considered as a Poisson process.*

Chapter 6

Continuous Time Markov Processes

6.1 Pure Birth Process

The simplest generalization of the Poisson process allows the probability of events to depend on how many have appeared previously, for example, the number of births goes up with the number of members in the population.

Consider a sequence $\{\lambda_k\}_{k \geq N}$ with $\lambda_k \geq 0$, $k \geq N$.

Postulates 6.1.1. *(Pure birth process)*

1) $\Pr(X(t+h) - X(t) = 1 \mid X(t) = j) = \lambda_j h + o(h)$ *for all* $j \in \mathbb{N}_0$*;*

2) $\Pr(X(t+h) - X(t) = 0 \mid X(t) = j) = 1 - \lambda_j h + o(h)$ *for all* $j \in \mathbb{N}_0$*;*

3) $X(0) = N$ *(initial population).*

Note that these postulates also hold for the Poisson process with $\lambda_k \equiv \lambda$.

Theorem 6.1.2. *Let* $P_n(t) = \Pr(N((0,t]) = n)$ *as above. Then the above postulates imply that*

$$P_n'(t) = -\lambda_n P_n(t) + \lambda_{n-1} P_{n-1}(t), \quad n \geq 1, \quad P_0'(t) = -\lambda_0 P_0(t)$$

with initial values

$$P_N(0) = 1, \quad P_n(0) = 0 \text{ for } n \neq N.$$

Proof. As for the Poisson process, first step analysis yields

$$P_n(t+h) = P_n(t)(1 - \lambda_n h + o(h)) + P_{n-1}(t)(\lambda_{n-1} h + o(h))$$

for $n \geq 1$ and

$$P_0(t+h) = P_0(t)(1 - \lambda_0 h + o(h))$$

which gives

$$\frac{P_n(t+h) - P_n(t)}{h} = -\lambda_n P_n(t) + P_{n-1}(t)\lambda_{n-1} + \frac{o(h)}{h},$$

for $n \geq 1$ and

$$\frac{P_0(t+h) - P_0(t)}{h} = -\lambda_0 P_0(t) + \frac{o(h)}{h}.$$

Letting $h \to 0$ we get

$$\frac{\mathrm{d}P_n(t)}{\mathrm{d}t} = -\lambda_n P_n(t) + \lambda_{n-1} P_{n-1}(t), \quad n \geq 1, \quad \text{and} \quad \frac{\mathrm{d}P_0(t)}{\mathrm{d}t} = -\lambda_0 P_0(t).$$

Note that
$$0 \equiv P_0(t) \equiv P_1(t) \equiv \ldots \equiv P_{N-1}(t).$$
Hence $P_N(t)$ satisfies
$$P_N'(t) = -\lambda_N P_N(t).$$
Since $P_N(0) = 1$ we get $P_N(t) = e^{-\lambda_N t}$.

Example 6.1.3. (Yule process) Consider a population of members who give birth, but do not die. Assume that each member acts independently and has a probability $\lambda h + o(h)$ of giving birth to a new member during a time interval of length h. Assume that the initial population at time $t = 0$ is $X(0) = 1$. If at time t the population has n members then

$$\begin{aligned} \Pr(X(t+h) - X(t) = 1 \mid X(t) = n) &= \binom{n}{1}(\lambda h + o(h))(1 - \lambda h + o(h))^{n-1} \\ &= n\lambda h(1 - O(h)) = n\lambda h + o(h). \end{aligned}$$

The probability of $k > 1$ members giving birth is

$$\binom{n}{k}(\lambda h + o(h))^k(1 - \lambda h + o(h))^{n-k} = o(h).$$

This is a special pure birth process with $\lambda_n = nh$ which we call a *Yule process.* The corresponding ODEs are
$$P_n'(t) = -n\lambda P_n(t) + (n-1)\lambda P_{n-1}(t), \quad P_1'(t) = -\lambda P_1(t)$$
with
$$P_1(1) = 1, \quad P_n(0) = 0 \text{ for } n \geq 1,$$
since the initial population is $X(0) = 1$. These differential equations can be solved iteratively to give
$$P_n(t) = e^{-\lambda t}(1 - e^{-\lambda t})^{n-1}.$$

6.2 Birth and Death Process

The pure birth process models early stages of population growth, but not population growth where members can die or be removed. So we generalize to Birth and Death Processes.

Postulates 6.2.1. *(Birth and Death Process) Consider non-negative sequences* $\{\lambda_k\}, \{\mu_k\}$, $k \in \mathbb{N}_0$.

1) $\Pr(X(t+h) - X(t) = 1 \mid X(t) = j) = \lambda_j h + o(h)$ *for all* $j \in \mathbb{N}_0$*;*

2) $\Pr(X(t+h) - X(t) = -1 \mid X(t) = j) = \mu_j h + o(h)$ *for all* $j \in \mathbb{N}_0$*;*

3) $\Pr(|X(t+h) - X(t)| > 1 \mid X(t) = j) = o(h)$ *for all* $j \in \mathbb{N}_0$.

Theorem 6.2.2. *Let* $P_n(t) = \Pr(N((0,t]) = n)$ *as above. Then the above postulates imply that*

$$\begin{aligned} P_n'(t) &= -(\lambda_n + \mu_n)P_n(t) + \lambda_{n-1}P_{n-1}(t) + \lambda_{n+1}P_{n+1}(t), \quad n \geq 1, \\ P_0'(t) &= -\lambda_0 P_0(t) + \mu_1 P_1(t). \end{aligned}$$

Proof. By first step analysis we have

$$\begin{aligned} P_n(t+h) &= P_n(t)(1-\lambda_n h - \mu_n h) + P_{n-1}(t)\lambda_{n-1}h + P_{n+1}(t)\mu_{n+1}h + o(h) \quad \text{for } n \in \mathbb{N}, \\ P_0(t+h) &= P_0(t)(1-\lambda_0 h) + P_1(t)\mu_1 h + o(h). \end{aligned}$$

The first term in the first equation is the probability that the population size equals n at time t and that no birth or death occurs in the time interval $(t, t+h]$. For $n \geq 1$ the second term is the probability that the population size equals $n-1$ at time t and one birth occurs in the time interval $(t, t+h]$. The third term is the probability that the population size equals $n+1$ at time t and one death occurs in the time interval $(t, t+h]$. Due to the postulate 3 the probability of more than one death or birth in the time interval $(t, t+h]$ is $o(h)$. Since the population size is a nonnegative number, the last term is absent in the case $n = 0$ and $\mu_0 = 0$. Rearranging we get

$$\begin{aligned} \frac{P_n(t+h) - P_n(t)}{h} &= -(\lambda_n + \mu_n)P_n(t) + \lambda_{n-1}P_{n-1}(t) + \mu_{n+1}P_{n+1}(t) + o(1) \\ \frac{P_0(t+h) - P_0(t)}{h} &= -\lambda_0 P_0(t) + \mu_1 P_1(t) + o(1). \end{aligned}$$

Letting $h \to 0$ we get the differential equations as required.