A Course in Stochastic Processes

José Manuel Corcuera
# Contents

1 Basic Concepts and Definitions ........................................... 3
  1.1 Definition of a Stochastic Process ............................... 3
  1.2 Examples .................................................................. 4
  1.3 Equivalence of Stochastic Processes ............................ 6
  1.4 Continuity Concepts .................................................. 6
  1.5 Main Classes of Random Processes ............................... 7
  1.6 Some Applications .................................................... 8
    1.6.1 Model of a Germination Process .......................... 8
    1.6.2 Modeling of Soil Erosion Effect ......................... 9
    1.6.3 Brownian motion .............................................. 9
  1.7 Problems .................................................................. 10

2 Discrete time martingales ...................................................... 13
  2.1 Conditional expectation .............................................. 13
  2.2 Discrete Time Martingales ......................................... 16
  2.3 Stopping times .......................................................... 19
  2.4 Optional stopping theorem ......................................... 21
  2.5 The Snell envelope and optimal stopping ...................... 22
  2.6 Decomposition of supermartingales ............................. 25
  2.7 Martingale convergence theorems ................................ 27

3 Markov Processes .............................................................. 31
  3.1 Discrete-time Markov chains ...................................... 31
    3.1.1 Class structure .................................................. 33
    3.1.2 Hitting times and absorption probabilities .......... 34
    3.1.3 Recurrence and transience .................................. 39
    3.1.4 Recurrence and transient in a random walk ....... 41
    3.1.5 Invariant distributions ....................................... 42
    3.1.6 Limiting behaviour ............................................ 47
    3.1.7 Ergodic theorem ................................................ 47
  3.2 Poisson process .......................................................... 48
    3.2.1 Exponential distribution ................................... 48
    3.2.2 Poisson process ................................................ 49
  3.3 Continuous-time Markov chains .................................. 52
Chapter 1

Basic Concepts and Definitions

1.1 Definition of a Stochastic Process

A stochastic process with state space $S$ is a collection of random variables \( \{X_t; t \in T\} \) defined on the same probability space \((\Omega, \mathcal{F}, P)\). The set $T$ is called its parameter set. If $T = \mathbb{N} = \{0, 1, 2, \ldots\}$, the process is said to be a discrete parameter process. If $T$ is not countable, the process is said to have a continuous parameter. In the latter case the usual examples are $T = \mathbb{R}_+ = [0, \infty)$ and $T = [a; b] \subset \mathbb{R}$. The index $t$ represents time, and then one thinks of $X_t$ as the state or the position of the process at time $t$. The state space is $\mathbb{R}$ in most usual examples, and then the process is said real-valued. There will be also examples where $S$ is $\mathbb{N}$, the set of all integers, or a finite set.

For every fixed $\omega \in \Omega$, the mapping

$$t \mapsto X_t(\omega)$$

defined on the parameter set $T$, is called a realization, trajectory, sample path or sample function of the process.

Let $\{X_t; t \in T\}$ be a real-valued stochastic process and $\{t_1 < \cdots < t_n\} \subset T$, then the probability distribution $P_{t_1, \ldots, t_n} = P \circ (X_{t_1}, \ldots, X_{t_n})^{-1}$ of the random vector

$$(X_{t_1}, \ldots, X_{t_n}) : \Omega \rightarrow \mathbb{R}^n$$

is called a finite-dimensional marginal distribution of the process $\{X_t; t \in T\}$. A theorem due to Kolmogorov, establishes the existence of a stochastic process associated with a given family of finite-dimensional distributions satisfying the consistence condition. Let $\Omega = \mathbb{R}^T$, the collection of all functions $\omega := (\omega(t))_{t \in T}$ from $T$ to $\mathbb{R}$. Define $X_t(\omega) = \omega(t)$. A set

$$C = \{\omega : X_{t_1}(\omega) \in B_1, \ldots, X_{t_n}(\omega) \in B_n\}$$

is called a finite-dimensional marginal distribution of the process $\{X_t; t \in T\}$. A theorem due to Kolmogorov, establishes the existence of a stochastic process associated with a given family of finite-dimensional distributions satisfying the consistence condition. Let $\Omega = \mathbb{R}^T$, the collection of all functions $\omega := (\omega(t))_{t \in T}$ from $T$ to $\mathbb{R}$. Define $X_t(\omega) = \omega(t)$. A set
for \( t_1 < \cdots < t_n \in T \) and \( B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R}^n) \) is called a cylinder set. Consider the \( \sigma \)-field \( \mathcal{F} \) generated by the cylinders sets.

**Theorem 1.1.1** Consider a family of probability measures

\[
\{P_{t_1, \ldots, t_n}, t_1 < \cdots < t_n, n \geq 1, t_i \in T\}
\]

such that: 1. \( P_{t_1, \ldots, t_n} \) is a probability on \( \mathbb{R}^n \). 2. (Consistence condition): If

\[
\{t_{k_1} < \cdots < t_{k_m}\} \subset \{t_1 < \cdots < t_n\},
\]

then \( P_{t_{k_1}, \ldots, t_{k_m}} \) is the marginal of \( P_{t_1, \ldots, t_n} \) corresponding to the indexes \( k_1, \ldots, k_m \).

Then, there exists a unique probability \( P \) on \( \mathcal{F} \) which has the family \( \{P_{t_1, \ldots, t_n}\} \) as finite-dimensional marginal distributions.

### 1.2 Examples

A real-valued process \( \{X_t, t \geq 0\} \) is called a second order process provided \( E(X^2) < \infty \) for all \( t \geq 0 \). The mean and the covariance function of a second order process \( \{X_t, t \geq 0\} \) are defined by

\[
m_X(t) = E(X_t),
\]

\[
\Gamma_X(s, t) = \text{Cov}(X_s; X_t) = E((X_s - m_X(s))(X_t - m_X(t))).
\]

The variance of the process \( \{X_t, t \geq 0\} \) is defined by

\[
\sigma_X^2(t) = \Gamma_X(t, t) = \text{Var}(X_t).
\]

**Example 1.2.1** Let \( X \) and \( Y \) be independent random variables. Consider the stochastic process with parameter \( t \in [0, \infty) \)

\[
X_t = tX + Y.
\]

The sample paths of this process are lines with random coefficients. The finite-dimensional marginal distributions are given by

\[
P(X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_n) = \int_{\mathbb{R}} F_X \left( \min_{1 \leq i \leq n} \frac{x_i - y}{t_i} \right) P_Y(dy).
\]

**Example 1.2.2** Consider the stochastic process

\[
X_t = A\cos(\varphi + \lambda t);
\]

where \( A \) and \( \varphi \) are independent random variables such that \( E(A) = 0, E(A^2) < \infty \) and \( \varphi \) is uniformly distributed on \([0, 2\pi]\). This is a second order process with

\[
m_X(t) = 0
\]

\[
\Gamma_X(s, t) = \frac{1}{2} E(A^2) \cos(\lambda (t - s)).
\]
1.2. EXAMPLES

Example 1.2.3 Arrival process: Consider the process of arrivals of customers at a store, and suppose the experiment is set up to measure the interarrival times. Suppose that the interarrival times are positive random variables \( X_1, X_2, \ldots \). Then, for each \( t \in [0; \infty) \), we put \( N_t = k \) if and only if the integer \( k \) is such that

\[
X_1 + \ldots + X_k \leq t < X_1 + \ldots + X_{k+1},
\]

and we put \( N_t = 0 \) if \( t < X_1 \). Then \( N_t \) is the number of arrivals in the time interval \([0, t]\). Notice that for each \( t \geq 0 \), \( N_t \) is a random variable taking values in the set \( S = \mathbb{N} \). Thus, \( \{N_t, t \geq 0\} \) is a continuous time process with values in the state space \( \mathbb{N} \). The sample paths of this process are non-decreasing, right continuous and they increase by jumps of size 1 at the times \( X_1 + \ldots + X_k \). On the other hand, \( N_t < \infty \) for all \( t \geq 0 \) if and only if

\[
\sum_{k=1}^{\infty} X_k = \infty.
\]

Example 1.2.4 Consider a discrete time stochastic process \( \{X_n, n = 0, 1, 2, \ldots\} \) with a finite number of states \( S = \{1, 2, 3\} \). The dynamics of the process is as follows. You move from state 1 to state 2 with probability \( \frac{1}{2} \). From state 3 you move either to 1 or to 2 with equal probability \( \frac{1}{2} \), and from 2 you jump to 3 with probability \( \frac{1}{3} \), otherwise stay at 2. This is an example of a Markov chain.

A real-valued stochastic process \( \{X_t, t \in T\} \) is said to be Gaussian or normal if its finite-dimensional marginal distributions are multi-dimensional Gaussian laws. The mean \( m_X(t) \) and the covariance function \( \Gamma_X(s, t) \) of a Gaussian process determine its finite-dimensional marginal distributions. Conversely, suppose that we are given an arbitrary function \( m : T \rightarrow \mathbb{R} \), and a symmetric function \( \Gamma : T \times T \rightarrow \mathbb{R} \), which is nonnegative definite, that is

\[
\sum_{i,j=1}^{n} \Gamma(t_i, t_j)a_i a_j \geq 0
\]

for all \( t_i \in T, a_i \in \mathbb{R} \), and \( n \geq 1 \). Then there exists a Gaussian process with mean \( m \) and covariance function \( \Gamma \).

Example 1.2.5 Let \( X \) and \( Y \) be random variables with joint Gaussian distribution. Then the process \( X_t = tX + Y, t \geq 0 \), is Gaussian with mean and covariance functions

\[
m_X(t) = tE(X) + E(Y);
\]

\[
\Gamma_X(s, t) = st \text{Var}(X) + (t + s) \text{Cov}(X, Y) + \text{Var}(Y).
\]

Example 1.2.6 Gaussian white noise: Consider a stochastic process \( \{X_t, t \in T\} \) such that the random variables \( X_t \) are independent and with the same law \( N(0, \sigma^2) \). Then, this process is Gaussian with mean and covariance functions,

\[
m_X(t) = 0
\]

\[
\Gamma_X(s, t) = \sigma^2 \delta_{st}.
\]
1.3 Equivalence of Stochastic Processes

**Definition 1.3.1** A stochastic process \( \{X_t, t \in T\} \) is equivalent to another stochastic process \( \{Y_t, t \in T\} \) if for each \( t \in T \)

\[
P\{X_t = Y_t\} = 1.
\]

We also say that \( \{X_t, t \in T\} \) is a version of \( \{Y_t, t \in T\} \). Two equivalent processes may have quite different sample paths.

**Example 1.3.1** Let \( \xi \) be a nonnegative random variable with continuous distribution function. Set \( T = [0, \infty) \). The processes

\[
X_t = 0 \\
Y_t = \begin{cases} 
0 & \text{if } \xi \neq t \\
1 & \text{if } \xi = t
\end{cases}
\]

are equivalent but their sample paths are different.

**Definition 1.3.2** Two stochastic processes \( \{X_t, t \in T\} \) and \( \{Y_t, t \in T\} \) are said to be indistinguishable if

\[
X_\omega = Y_\omega \quad \text{for all } \omega \notin N, \quad P(N) = 0.
\]

Two stochastic processes which have right continuous sample paths and are equivalent, then they are indistinguishable.

Two discrete time stochastic processes which are equivalent, they are also indistinguishable.

1.4 Continuity Concepts

**Definition 1.4.1** A real-valued stochastic process \( \{X_t, t \in T\} \), where \( T \) is an interval of \( \mathbb{R} \), is said to be continuous in probability if, for any \( \varepsilon > 0 \) and every \( t \in T \)

\[
\lim_{s \to t} P(|X_t - X_s| > \varepsilon) = 0.
\]

**Definition 1.4.2** Fix \( p \geq 1 \). Let \( \{X_t, t \in T\} \) be a real-valued stochastic process, where \( T \) is an interval of \( \mathbb{R} \), such that \( E(|X_t|^p) < \infty \), for all \( t \in T \). The process \( \{X_t, t \in T\} \) is said to be continuous in mean of order \( p \) if

\[
\lim_{s \to t} E(|X_t - X_s|^p) = 0.
\]

Continuity in mean of order \( p \) implies continuity in probability. However, the continuity in probability (or in mean of order \( p \)) does not necessarily implies that the sample paths of the process are continuous.

In order to show that a given stochastic process have continuous sample paths it is enough to have suitable estimations on the moments of the increments of the process. The following continuity criterion by Kolmogorov provides a sufficient condition of this type:
1.5. MAIN CLASSES OF RANDOM PROCESSES

Proposition 1.4.1 (Kolmogorov continuity criterion) Let \( \{X_t, t \in T\} \) be a real-valued stochastic process and \( T \) is a finite interval. Suppose that there exist constants \( \alpha > 1 \) and \( p > 0 \) such that

\[
E(|X_t - X_s|^p) \leq c_T |t - s|^\alpha \tag{1.1}
\]

for all \( s, t \in T \). Then, there exists a version of the process \( \{X_t, t \in T\} \) with continuous sample paths.

Condition (1.1) also provides some information about the modulus of continuity of the sample paths of the process:

\[
m_T(\omega, \delta) := \sup_{s, t \in T, |t - s| < \delta} |X_t(\omega) - X_s(\omega)|.
\]

In particular if \( T = [a, b] \), for each \( \varepsilon > 0 \) there exists a random variable \( G_\varepsilon \) such that, with probability one,

\[
m_T(\omega, \delta) \leq G_\varepsilon |b - a|^{\frac{\alpha - 1}{p}} - \varepsilon.
\]

Moreover, \( E(G_\varepsilon^p) < \infty \).

1.5 Main Classes of Random Processes

1. Processes with independent increments. Let \( X := \{X_t, t \in T\} \) be a real-valued stochastic process on a probability space \((\Omega, \mathcal{F}, P)\), where \( T \subset \mathbb{R} \) is an interval.

**Definition 1.5.1** The process \( X \) is said to have independent increments if for any finite subset \( \{t_0 < \cdots < t_n\} \subset T \), the increments

\[
X_{t_0}, X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}
\]

are independent random variables.

**Remark 1.5.1** If \( T = [0, \infty) \), from the definition, it follows that all the marginal distributions are determined by \( X_0 \) and \( X_t - X_s, s < t \in T \). The most important processes with independent increments are the Poisson process and the Wiener process (or Brownian motion).


**Definition 1.5.2** The process \( X \) is said to be a Markov process with a state space \((\mathbb{R}, \mathcal{B}(\mathbb{R}))\) if for any \( \{t_1 < \cdots < t_n\} \subset T \) and \( B \in \mathcal{B}(\mathbb{R}) \),

\[
P(X_{t_n} \in B | X_{t_1}, \ldots, X_{t_{n-1}}) = P(X_{t_n} \in B | X_{t_{n-1}}) \quad (a.s.) \tag{1.2}
\]

**Remark 1.5.2** Property (2.3) is called the Markov property.

Definition 1.5.3 The process $X$ is said to be strictly stationary if for any \( \{ t_1 < \cdots < t_n \} \subset T \) and \( \{ t_1 + h < \cdots < t_n + h \} \subset T \) any \( (B_k)_{1 \leq k \leq n} \in B(\mathbb{R}) \)

\[
P(X_{t_1+h} \in B_1, \ldots, X_{t_n+h} \in B_n) = P(X_{t_1} \in B_1, \ldots, X_{t_n} \in B_n)
\]

4. Wide sense stationary processes.

Definition 1.5.4 A second order process $X$ is said to be wide sense stationary if for all $t \in T$ and $t+h \in T$

\[
E(X_{t+h}) = E(X_t)
\]

and

\[
Cov(X_{t+h}, X_{s+h}) = Cov(X_t, X_s).
\]

5. Martingales.

Definition 1.5.5 A process $X$, such that $E(|X|) < \infty$ is said to be a martingale if, for every \( \{ t_1 < \cdots < t_n \} \subset T \),

\[
E(X_{t_n}|X_{t_1}, \ldots, X_{t_{n-1}}) = X_{t_{n-1}} \text{ (a.s.)}.
\]

If

\[
E(X_{t_n}|X_{t_1}, \ldots, X_{t_{n-1}}) \leq X_{t_{n-1}} \text{ (a.s.),}
\]

the process $X$ is a supermartingale. Finally if

\[
E(X_{t_n}|X_{t_1}, \ldots, X_{t_{n-1}}) \geq X_{t_{n-1}} \text{ (a.s.),}
\]

it is said to be a submartingale.

1.6 Some Applications

1.6.1 Model of a Germination Process

Let $p$ be the probability that a seed may fall to germinate. If there is germination assume that the germination time, $T$, follows an exponential distribution with parameter $\lambda$. Then

\[
P(T \leq t) = P(T \leq t|\text{germination})P(\text{germination})
\]

\[
= (1 - e^{-\lambda t})(1 - p).
\]

If we have $n$ plants and $X_t$ is the number of them that have germinated, then

\[
X_t = \sum_{i=1}^{n} 1\{T_i \leq t\},
\]

where $T_i$ is the germination time of the $i$-plant. If we assume that $\{T_i\}_{i=1}^{N}$ is an i.i.d. sequence, then

\[
X_t \sim \text{Bi} \left( N, (1 - e^{-\lambda t})(1 - p) \right).
\]
1.6.2 Modeling of Soil Erosion Effect

Let $Y_1, Y_2, ..., Y_n, ...$ be the sequence of annual yields of a given crop in an area without erosion and $Z_1, Z_2, ... Z_n, ...$ a sequence of factors with common support in $(0, 1]$. Then we can model the crop yield in year $n$ as

$$X_n = Y_n \prod_{i=1}^{n} Z_i,$$

and to assume that $Y$ and $Z$ are, respectively, sequences of i.i.d random variables and that $Y$ and $Z$ are independent.

1.6.3 Brownian motion

In 1827 Robert Brown noticed the irregular but ceaseless motion of pollen particles suspended in a liquid. Assume that $\xi_t$ is the coordinate, say $x$, of the position of a pollen particle. Then $(\xi_t)$ is a continuos-time process. If we assume that $\xi_t - \xi_s$ has a distribution depending on $t - s$, independent of what happened before $s$ and that $(\xi_t)$ is a continuous process then it can be proved that

$$\xi_t \sim N(\mu t, \sigma^2 t),$$

where $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}_+$. If $\mu = 0$ and $\sigma = 1$ we have the so called standard Brownian motion. Since

$$(\xi_t - \xi_s)^2 = \xi_t^2 + \xi_s^2 - 2\xi_t \xi_s,$$

we have that

$$|t - s| = E((\xi_t - \xi_s)^2) = E(\xi_t^2) + E(\xi_s^2) - 2\text{Cov}(\xi_t, \xi_s),$$

and

$$\text{Cov}(\xi_t, \xi_s) = \frac{1}{2}(t + s - |t - s|) = \min(s, t).$$
1.7 Problemes

1. Sigui \( \Omega = [0, 1] \), \( \mathcal{F} = \mathcal{B}([0, 1]) \) i \( P \) la mesura de probabilitat uniforme en \( \Omega \). Considereu el procés \( X_t(\omega) = t\omega, \ t \in [0, 1] \). Trobeu les trajectòries i les distribucions en una i dos dimensions.

2. Sigui \( X_t = A \cos \alpha t + B \sin \alpha t \) i \( Y_t = R \cos(\alpha t + \Theta) \). On \( \alpha > 0 \), \( A \) i \( B \) són variables aleatòries independents amb distribució \( N(0, \sigma^2) \). \( R \) i \( \Theta \) són també variables aleatòries independents on \( \Theta \sim \text{Unif}(0, 2\pi) \) i \( R \) té densitat

\[
 f_R(r) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right) 1(0, +\infty)(x).
\]

Demostreu que els processos \( \{X_t\}_{t \geq 0} \) i \( \{Y_t\}_{t \geq 0} \) tenen la mateixa llei.

3. Sigui \( X_1 \) i \( X_2 \) variables aleatòries independents definides en el mateix espai de probabilitat \( (\Omega, \mathcal{F}, P) \) i amb la mateixa distribució normal estàndard. Sigui \( \{Y_t\}_{t \geq 0} \) el procés estocàstic definit per

\[
 Y_t = (X_1 + X_2) t.
\]

Determineu les distribucions en dimensió finita del procés. Si \( A \) és el conjunt de les trajectòries no negatives del procés, calculeu \( P(A) \).

4. Sigui \( \{Y_t\}_{t \geq 0} \) el procés estocàstic definit per

\[
 Y_t = X + \alpha t, \quad \alpha > 1,
\]

on \( X \) és una variable aleatòria amb llei \( N(0,1) \). Sigui \( D \subset [0, +\infty) \) un conjunt finit o infinit numerable. Determineu:

(a) \( P(Y_t = 0, \text{per almenys un } t \in D) \),

(b) \( P(Y_t = 0, \text{per almenys un } t \in [1, 2]) \).

5. Sigui \( X \) i \( Y \) variables aleatòries definides en \( (\Omega, \mathcal{F}, P) \), on \( Y \sim N(0, 1) \). Sigui \( \{Z_t\}_{t \geq 0} \) el següent procés estocàstic

\[
 Z_t = X + t(Y + t).
\]

Sigui \( A \) el conjunt de trajectòries no decreixents d’aquest procés. Determineu \( P(A) \).

6. Sigui \( \{X_i\}_{i=1,...,n} \) i \( \{Y_i\}_{i=1,...,n} \) variables aleatòries tal que \( \mathbb{E}[X_i] = \mathbb{E}[Y_i] = 0 \), \( \text{Var}[X_i] = \text{Var}[Y_i] = \sigma_i < +\infty \) i \( \mathbb{E}[X_i X_j] = \mathbb{E}[Y_i Y_j] = \mathbb{E}[X_i Y_j] = 0 \) per a tot \( i \neq j \). Sigui \( \{Z_t\}_{t \geq 0} \) el procés estocàstic definit per

\[
 Z_t = \sum_{i=1}^{n} \{X_i \cos(\lambda_i t) + Y_i \sin(\lambda_i t)\}.
\]

Determineu la seva funció de covariància. És \( \{Z_t\}_{t \geq 0} \) estacionari en sentit ampli?
7. Diem que un procés \( \{N_t\}_{t \geq 0} \) és un procés de Poisson homogeni de paràmetre \( \lambda > 0 \), si té increments independents, \( N_t = 0 \) i per a tot \( 0 < t_1 < t_2 \) es compleix

\[
P(N_{t_2} - N_{t_1} = k) = \frac{(\lambda (t_2 - t_1))^k}{k!} e^{-\lambda (t_2 - t_1)}, \quad k \in \mathbb{Z}_+.
\]

Sigui \( \{X_k\}_{k \geq 0} \) una successió de v.a. i.i.d. amb \( \mathbb{E}[X_k] = 0 \) i \( \text{Var}[X_k] = \sigma^2 < +\infty \). Sigui \( \{N_t\}_{t \geq 0} \) un procés de Poisson homogeni de parametre \( \lambda > 0 \), independent de \( \{X_k\}_{k \geq 0} \). És el procés \( \{Y_t\}_{t \geq 0} \), definit per \( Y_t = X_{N_t} \), estacionari en sentit ampli? I estrictament estacionari?

8. Sigui \( \{Y_t\}_{t \geq 0} \) el procés estocàstic definit per

\[
Y_t = \alpha \sin (\beta t + X),
\]

on \( \alpha, \beta > 0 \) i \( X \sim \mathcal{N}(0,1) \). Calculeu \( \mathbb{E}[Y_t] \) i \( \mathbb{E}[Y_s Y_t] \). És el procés \( \{Y_t\}_{t \geq 0} \) estacionari en sentit ampli?
Chapter 2

Discrete time martingales

We will first introduce the notion of conditional expectation of a random variable $X$ with respect to a $\sigma$-field $\mathcal{B} \subset \mathcal{F}$ in a probability space $(\Omega, \mathcal{F}, P)$.

2.1 Conditional expectation

Consider an integrable random variable $X$ defined in a probability space $(\Omega, \mathcal{F}, P)$, and $\sigma$-field $\mathcal{B} \subset \mathcal{F}$. We define the conditional expectation of $X$ given $\mathcal{B}$ (denoted by $E(X|\mathcal{B})$) to be any integrable random variable $Z$ that satisfies the following two properties:

(i) $Z$ is measurable with respect to $\mathcal{B}$.

(ii) For all $A \in \mathcal{B}$, $E(Z1_A) = E(X1_A)$

It can be proved that there exists a unique (up to modifications on sets of probability zero) random variable satisfying these properties. That is, if $\tilde{Z}$ and $Z$ satisfy the above properties, then $Z = \tilde{Z}$, $P$-almost surely.

Property (ii) implies that for any bounded and $\mathcal{B}$-measurable random variable $Y$ we have

$$E(E(X|\mathcal{B})Y) = E(XY)$$ (2.1)

Example 2.1.1 Consider the particular case where the $\sigma$-field $\mathcal{B}$ is generated by a finite partition $\{B_1, ..., B_m\}$. In this case, the conditional expectation $E(X|\mathcal{B})$ is a discrete random variable that takes the constant value $E(X|B_j)$ on each set $B_j$:

$$E(X|\mathcal{B}) = \sum_{j=1}^{m} \frac{E(X1_{B_j})}{P(B_j)}1_{B_j}.$$ 

Here are some properties of the conditional expectations in the general case:

1. The conditional expectation is lineal: $E(aX+bY|\mathcal{B}) = aE(X|\mathcal{B})+bE(Y|\mathcal{B})$. 

2. A random variable and its conditional expectation have the same expectation: $E(E(X|\mathcal{B})) = E(X)$. This follows from property (ii) taking $A = \Omega$.

3. If $X$ and $\mathcal{B}$ are independent, then $E(X|\mathcal{B}) = E(X)$. In fact, the constant $E(X)$ is clearly $\mathcal{B}$-measurable, and for all $A \in \mathcal{B}$ we have $E(X 1_A) = E(E(X 1_A)) = E(E(X) 1_A) = E(E(X) 1_A)$.

4. If $X$ is $\mathcal{B}$-measurable, then $E(E(X|\mathcal{B})) = X$.

5. If $Y$ is a bounded and $\mathcal{B}$-measurable random variable, then $E(Y X|\mathcal{B}) = Y E(X|\mathcal{B})$. In fact, the constant $E(X)$ is clearly $\mathcal{B}$-measurable, and for all $A \in \mathcal{B}$ we have $E(Y X 1_A) = E(Y E(X) 1_A) = E(Y E(X) 1_A)$, where the equality follows from (2.1). This property means that $\mathcal{B}$-measurable random variables behave as constants and can be factorized out of the conditional expectation with respect to $\mathcal{B}$. This property holds if $X, Y \in L^2(\Omega)$.

6. Given two $\sigma$-fields $\mathcal{C} \subset \mathcal{B}$, then $E(E(X|\mathcal{B})|\mathcal{C}) = E(X|\mathcal{C})$.

7. Consider two random variable $X$ and $Z$, such that $Z$ is $\mathcal{B}$-measurable and $X$ is independent of $\mathcal{B}$. Consider a measurable function $h(x, z)$ such that the composition $h(X, Z)$ is an integrable random variable. Then, we have $E(h(X, Z)|\mathcal{B}) = E(h(X; z)|z = Z)$. That is, we first compute the expectation $E(h(X; z))$ for any fixed value $z$ of the random variable $Z$ and, afterwards, we replace $z$ by $Z$.

Conditional expectation has properties similar to those of ordinary expectation. For instance, the following monotone property holds:

$$X \leq Y \Rightarrow E(X|\mathcal{B}) \leq E(Y|\mathcal{B}).$$

This implies $|E(X|\mathcal{B})| \leq E(|X||\mathcal{B})$.

Jensen’s inequality also holds. That is, if $\varphi$ is a convex function defined in an open interval, such that $E(|\varphi(X)|) < \infty$, then

$$\varphi(E(X|\mathcal{B})) \leq E(\varphi(X)|\mathcal{B}).$$

(2.2)

In particular, if we take $\varphi(x) = |x|^p$ with $p \geq 1$, and $E(|X|^p) < \infty$, we obtain

$$|E(X|\mathcal{B})|^p \leq E(|X|^p|\mathcal{B}),$$

hence, taking expectations,

$$E(|E(X|\mathcal{B})|^p) \leq E(|X|^p).$$

(2.3)

We can define the conditional probability of an event $C \in \mathcal{F}$ given a $\sigma$-field $\mathcal{B}$ as

$$P(C|\mathcal{B}) = E(1_C|\mathcal{B}).$$
Suppose that the $\sigma$-field $B$ is generated by a finite collection of random variables $Y_1, ..., Y_m$. In this case, we will denote the conditional expectation of $X$ given $B$ by $E(X|Y_1, ..., Y_m)$ and this conditional expectation is the mean of the conditional distribution of $X$ given $Y_1, ..., Y_m$. The conditional distribution of $X$ given $Y_1, ..., Y_m$ is a family of distributions $p(dx|y_1, ..., y_m)$ parametrized by the possible values $y_1, ..., y_m$ of the random variables $Y_1, ..., Y_m$, such that for all $a < b$

$$P(a \leq X \leq b|Y_1, ..., Y_m) = \int_a^b p(dx|Y_1, ..., Y_m).$$

Then, this implies that

$$E(X|Y_1, ..., Y_m) = \int_{\mathbb{R}} xp(dx|Y_1, ..., Y_m).$$

Notice that the conditional expectation $E(X|Y_1, ..., Y_m)$ is a function $g(Y_1, ..., Y_m)$ of the variables $Y_1, ..., Y_m$, where

$$g(y_1, ..., y_m) = \int_{\mathbb{R}} xp(dx|y_1, ..., y_m).$$

In particular, if the random variables $X, Y_1, ..., Y_m$ have a joint density $f(x, y_1, ..., y_m)$, then the conditional distribution has the density:

$$f(x|y_1, ..., y_m) = \frac{f(x, y_1, ..., y_m)}{\int_{-\infty}^{\infty} f(x, y_1, ..., y_m)dx},$$

and

$$E(X|Y_1, ..., Y_m) = \int_{-\infty}^{\infty} x f(x|Y_1, ..., Y_m)dx.$$

The set of all square integrable random variables, denoted by $L^2$, is a Hilbert space with the scalar product

$$\langle Z, Y \rangle = E(ZY).$$

Then, the set of square integrable and $B$-measurable random variables, denoted by $L^2(\Omega, B, P)$ is a closed subspace of $L^2(\Omega, F, P)$. Then, given a random variable $X$ such that $E(X^2) < \infty$, the conditional expectation $E(X|B)$ is the projection of $X$ on the subspace $L^2(\Omega, B, P)$. In fact, we have:

(i) $E(X|B)$ belongs to $L^2(\Omega, B, P)$ because it is a $B$-measurable random variable and it is square integrable due to (2.3).

$X - E(X|B)$ is orthogonal to the subspace $L^2(\Omega, B, P)$. In fact, for all $Z \in L^2(\Omega, B, P)$ we have, using the property 5,

$$E[(X - E(X|B))Z] = E(XZ) - E(E(X|B)Z)$$

$$= E(XZ) - E(E(XZ|B)) = 0.$$
CHAPTER 2. DISCRETE TIME MARTINGALES

As a consequence, $E(X|\mathcal{B})$ is the random variable in $L^2(\Omega, \mathcal{B}, P)$ that minimizes the mean square error:

$$E[(X - E(X|\mathcal{B}))^2] = \min_{Y \in L^2(\Omega, \mathcal{B}, P)} E[(X - Y)^2].$$

This follows from the relation

$$E[(X - Y)^2] = E[(X - E(X|\mathcal{B}))^2] + E[(E(X|\mathcal{B}) - Y)^2],$$

and it means that the conditional expectation is the optimal estimator of $X$ given the $\sigma$-field $\mathcal{B}$.

2.2 Discrete Time Martingales

In this section we consider a probability space $(\Omega, \mathcal{F}, P)$ and a nondecreasing sequence of $\sigma$-fields

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots \subseteq \mathcal{F}_n \subseteq \ldots$$

contained in $\mathcal{F}$. $\mathcal{F} := \{\mathcal{F}_n\}_{n \geq 0}$ is called a filtration. A sequence of real random variables $M = \{M_n\}_{n \geq 0}$ is called a martingale with respect to the filtration $\mathcal{F}$ if:

(i) For each $n \geq 0$, $M_n$ is $\mathcal{F}_n$-measurable (that is, $M$ is adapted to the filtration $\mathcal{F}$).

(ii) For each $n \geq 0$, $E(|M_n|) < \infty$.

(iii) For each $n \geq 0$,

$$E(M_{n+1}|\mathcal{F}_n) = M_n.$$

The sequence $M = \{M_n\}_{n \geq 0}$ is called a supermartingale (or submartingale) if property (iii) is replaced by $E(M_{n+1}|\mathcal{F}_n) \leq M_n$ (or $E(M_{n+1}|\mathcal{F}_n) \geq M_n$).

Notice that the martingale property implies that $E(M_n) = E(M_0)$ for all $n \geq 0$.

On the other hand, condition (iii) can also be written as

$$E(\Delta M_n|\mathcal{F}_{n-1}) = 0,$$

for all $n \geq 1$, where $\Delta M_n = M_n - M_{n-1}$.

**Example 2.2.1** Suppose that $\{\xi_n, n \geq 1\}$ are i.i.d. centered random variables. Set $M_0 = 0$ and $M_n = \xi_1 + \ldots + \xi_n$, for $n \geq 1$. Then $M_n$ is a martingale w.r.t $(\mathcal{F}_n)$, with $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$ for $n \geq 1$, and $\mathcal{F}_0 = \{\phi, \Omega\}$. In fact,

**Example 2.2.2** Suppose that $\{\xi_n, n \geq 1\}$ are i.i.d. random variables such that $P(\xi_n = 1) = 1 - P(\xi_n = -1) = p$, on $0 < p < 1$. Then $M_n = \left(\frac{1-p}{p}\right)^{\xi_1 + \ldots + \xi_n}$.
2.2. DISCRETE TIME MARTINGALES

$M_0 = 1$ is a martingale w.r.t $(\mathcal{F}_n)$, with $\mathcal{F}_n = \sigma(\xi_1, ..., \xi_n)$ for $n \geq 1$, and $\mathcal{F}_0 = \{\phi, \Omega\}$. In fact,

$$E(M_{n+1}|\mathcal{F}_n) = E\left(\left(\frac{1-p}{p}\right)^{\xi_1+...+\xi_{n+1}}|\mathcal{F}_n\right)$$

$$= \left(\frac{1-p}{p}\right)^{\xi_1+...+\xi_n} E\left(\left(\frac{1-p}{p}\right)^{\xi_{n+1}}|\mathcal{F}_n\right)$$

$$= M_n E\left(\left(\frac{1-p}{p}\right)^{\xi_{n+1}}\right) = M_n.$$

In the two previous examples, $\mathcal{F}_n = \sigma(M_0, \ldots, M_n)$, for all $n > 0$. That is, $(\mathcal{F}_n)$ is the filtration generated by the process $(M_n)$. Usually, when the filtration is not mentioned, we will take $\mathcal{F}_n = \sigma(M_0, \ldots, M_n)$, for all $n \geq 0$. This is always possible due to the following result:

**Proposition 2.2.1** Suppose $(M_n)_{n \geq 0}$ is a martingale with respect to a filtration $(\mathcal{G}_n)$. Let $\mathcal{F}_n = \sigma(M_0, \ldots, M_n) \subseteq \mathcal{G}_n$. Then $(M_n)_{n \geq 0}$ is a martingale with respect to the filtration $(\mathcal{F}_n)$.

**Proof.** We have, by the properties 6 and 4 of the conditional expectation,

$$E(M_{n+1}|\mathcal{F}_n) = E(E(M_{n+1}|\mathcal{G}_n)|\mathcal{F}_n) = E(M_n|\mathcal{F}_n) = M_n.$$

---

Some elementary properties of martingales:
1. If $(M_n)$ is a martingale, then for all $m \geq n$ we have

$$E(M_m|\mathcal{F}_n) = M_n.$$

In fact, for $m > n + 1$

$$E(M_m|\mathcal{F}_n) = E(E(M_m|\mathcal{F}_{m-1})|\mathcal{F}_n)$$

$$= E(M_{m-1}|\mathcal{F}_n) = \ldots = E(M_{n+1}|\mathcal{F}_n) = M_n.$$

2. $(M_n)$ is a submartingale if and only if $-M_n$ is a supermartingale.
3. If $(M_n)$ is a martingale and $\varphi$ is a convex function, defined in an open interval, such that $E(|\varphi(M_n)|) < \infty$ for all $n \geq 0$, then $\varphi(M_n)$ is a submartingale. In fact, by Jensen’s inequality for the conditional expectation we have

$$E(\varphi(M_{n+1})|\mathcal{F}_n) \geq \varphi(E(M_{n+1}|\mathcal{F}_n)) = \varphi(M_n).$$

In particular, if $(M_n)$ is a martingale such that $E(|M_n|^p) < \infty$ for all $n \geq 0$ and for some $p \geq 1$, then $(|M_n|^p)$ is a submartingale.
4. If \((M_n)\) is a submartingale and \(\varphi\) is a convex an increasing function such that \(E(|\varphi(M_n)|) < \infty\) for all \(n \geq 0\), then \((\varphi(M_n))\) is a submartingale. In fact, by Jensen’s inequality for the conditional expectation we have

\[
E(\varphi(M_{n+1})|\mathcal{F}_n) \geq \varphi(E(M_{n+1}|\mathcal{F}_n)) \geq \varphi(M_n).
\]

In particular, if \((M_n)\) is a submartingale, then \((M_n^+)\) and \((M_n \vee a)\) are submartingales.

Suppose that \((\mathcal{F}_n)_{n \geq 0}\) is a given filtration. We say that \((H_n)_{n \geq 1}\) is a predictable sequence of random variables if for each \(n \geq 1\), \(H_n\) is \(\mathcal{F}_{n-1}\)-measurable.

We define the martingale transform of \((M_n)\) by \((H \cdot M)_n\) as the sequence

\[
(H \cdot M)_n = M_0 + \sum_{j=1}^{n} H_j \Delta M_j, n \geq 1.
\]

**Proposition 2.2.2** If \((M_n)_{n \geq 0}\) is a (sub)martingale and \((H_n)_{n \geq 1}\) is a bounded (nonnegative) predictable sequence, then the martingale transform \((H \cdot M)_n\) is a (sub)martingale.

**Proof.** Clearly, for each \(n \geq 0\) the random variable \((H \cdot M)_n\) is \(\mathcal{F}_n\)-measurable and integrable. On the other hand, if \(n \geq 0\) we have

\[
E(\Delta(H \cdot M)_{n+1}|\mathcal{F}_n) = E(H_{n+1}\Delta M_{n+1}|\mathcal{F}_n) = H_{n+1}E(\Delta M_{n+1}|\mathcal{F}_n) = 0.
\]

**Remark 2.2.1** We may think of \(H_n\) as the amount of money a gambler will bet at time \(n\). Suppose that \(\Delta M_n = M_n - M_{n-1}\) is the amount of money a gambler can win or lose at every step of the game if the bet is 1 Euro, and \(M_0\) is the initial capital of the gambler. Then, \(M_n\) will be the fortune of the gambler at time \(n\) if in each step the bet is 1 Euro, and \((H \cdot M)_n\) will be the fortune of the gambler as he uses the gambling system \((H_n)\). The fact that \((M_n)\) is a martingale means that the game is fair. So, the previous proposition tells us that if a game is fair, it is also fair regardless the gambling system \((H_n)\).

**Example 2.2.3** Suppose that \(M_0 = 0\) and \(M_n = \xi_1 + \ldots + \xi_n, n \geq 1\), where \(\{\xi_n, n \geq 1\}\) are i.i.d. random variables such that \(P(\xi_n = 1) = P(\xi_n = -1) = \frac{1}{2}\). \((M_n)\) is a martingale. A famous system to bet is "the doubling strategy" defined as

\[
H_1 = 1
\]

\[
H_n = \begin{cases} 
2^{n-1} & \text{if } \xi_1 = \ldots = \xi_{n-1} = -1 \\
0 & \text{otherwise}
\end{cases}, n > 1,
\]
then, if we stop at $\tau = \inf\{k, \xi_k = 1\}$ we will have

$$(H \cdot M)_{\tau} = \sum_{i=1}^{\tau} H_n \Delta M_n$$

$$= - \sum_{i=1}^{\tau-1} 2^{n-1} + 2^n = 1! (a.s)$$

and our initial capital was zero! Note that $P(\tau = k) = (\frac{1}{2})^k$ and so $P(\tau < \infty) = 1$ then we will stop at a finite time. However, that result does not contradict our previous proposition, since if we stop the si process by $\tau$ and we look what happens at a fix time $n$ and its relation with previous times we obtain a martingale, that is

$$E((H \cdot M)_{n \wedge \tau} | \mathcal{F}_{n-1}) = (H \cdot M)_{n-1 \wedge \tau}$$ for all $n \geq 1$.

It is remarkable that the term martingale is originated from a French acronym for the gambling strategy described above.

### 2.3 Stopping times

Consider a filtration $\mathbb{F} = \{\mathcal{F}_n\}_{n \geq 0}$ in a probability space $(\Omega, \mathcal{F}, P)$. A (generalized) random variable is called a stopping time if $T$ indicates the time when an observable event happens. In this way at time $n$ we know the value of $T \wedge (n+1)$, in other words $T \wedge (n+1)$ is $\mathcal{F}_n$-medible, so equivalently $\{T \leq n\} \in \mathcal{F}_n$ for all $n \geq 0$.

**Definition 2.3.1** A (generalized) random variable $T$ is called a stopping time if

$$T : \Omega \rightarrow \mathbb{Z}_+ \cup \{\infty\},$$

satisfies $\{T \leq n\} \in \mathcal{F}_n$.

**Proposition 2.3.1** $T$ is an $\mathbb{F}$-stopping time iff

$$\{T = n\} \in \mathcal{F}_n, \text{ for all } n \geq 0.$$  

**Proof.**

$$\{T \leq n\} = \bigcup_{j=1}^{n} \{T = j\}$$

$$\{T = n\} = \{T \leq n\} \cap \{T \leq n-1\}^c.$$  

**Example 2.3.1** Consider a discrete time stochastic process $(X_n)$ adapted to $\mathbb{F}$. Let $A$ be a subset of the space state. Then the first hitting time of $A$, $T_A$ is a stopping time because

$$\{T_A = n\} = \{X_0 \notin A, X_1 \notin A, ..., X_{n-1} \notin A, X_n \in A\}.$$  

**Remark 2.3.1** We could consider random times with certain delay, for instance if $\{T \leq n - 1\} \in \mathcal{F}_n$. Or even we can consider predictable times: $\{T \leq n\} \in \mathcal{F}_{n-1}$. 
Remark 2.3.2  The extension of the notion of stopping time to continuous time
is evident: the (generalized) random variable
\[ T : \Omega \to [0, \infty] \]
is a stopping time if \{T \leq t\} \in F_t. Random times with infinitesimal delay are
called optional times: \{T < t\} \in F_t (equivalently \{T \leq t\} \in F_{t+} := \cap_{s>t} F_s),
random times with infinitesimal anticipation are called predictable times: \{T \leq t\} \in F_{t-} := \sigma(\cup_{s<t} F_s).

Consider a real-valued stochastic process \{X_t, t \geq 0\} with continuous tra-
jectories, and adapted to \( F = \{F_t, t \geq 0\} \). Assume \( X_0 = 0 \) and \( a > 0 \). The first
passage time for a level \( a \in \mathbb{R} \) defined by
\[ T_a := \inf \{t > 0 : X_t = a\} \]
is a stopping time because
\[ \{T_a \leq t\} = \{ \sup_{0 \leq s \leq t} X_s \geq a\} = \{ \sup_{0 \leq s \leq t, s \in \mathbb{Q}} X_s \geq a\} \in F_t. \]

Properties of stopping times:
1. If \( S \) and \( T \) are stopping times, so are \( S \lor T \) and \( S \land T \). In fact, this a
   consequence of the relations
   \[ \{S \lor T \leq t\} = \{S \leq t\} \cap \{T \leq t\}, \]
   \[ \{S \land T \leq t\} = \{S \leq t\} \cup \{T \leq t\}. \]

2. Given a stopping time \( T \), we can define the \( \sigma \)-field
   \[ F_T = \{A : A \cap \{T \leq t\} \in F_t, \text{ for all } t \geq 0\}. \]
   \( F_T \) is a \( \sigma \)-field because it contains the empty set, it is stable by complements
due to
   \[ A^c \cap \{T \leq t\} = (A \cup \{T > t\})^c = ((A \cap \{T \leq t\}) \cup \{T > t\})^c, \]
   and it is stable by countable intersections.

3. If \( S \) and \( T \) are stopping times such that \( S \leq T \), then \( F_S \subset F_T \). In fact, if \( A \in F_S \), then
   \[ A \cap \{T \leq t\} = A \cap \{S \leq t\} \cap \{T \leq t\} \in F_t \]
   for all \( t \geq 0 \).

Remark 2.3.3  The \( \sigma \)-field \( F_T \) have the meaning of the information before the
stopping time \( T \). We observe what happens until time \( T \), time \( T \) included.
2.4. OPTIONAL STOPPING THEOREM

4. Let \( \{X_t\} \) be an adapted stochastic process (with discrete or continuous parameter) and let \( T \) be a stopping time. If the parameter is continuous, we assume that the trajectories of the process \( \{X_t\} \) are right-continuous. Then the random variable
\[
X_T(\omega) = X_{T(\omega)}(\omega)
\]
is \( \mathcal{F}_T \)-measurable. In discrete time this property follows from the relation
\[
\{X_T \in B\} \cap \{T = n\} = \{X_n \in B\} \cap \{T = n\} \in \mathcal{F}_n
\]
for any subset \( B \) of the state space (Borel set if the state space is \( \mathbb{R} \)).

2.4 Optional stopping theorem

Consider a discrete time filtration and suppose that \( T \) is a stopping time. Then, the process
\[
H_n = 1_{\{T \geq n\}}
\]
is predictable. In fact, \( \{T \geq n\} = \{T \leq n - 1\}^c \in \mathcal{F}_{n-1} \). The martingale transform of \( M_n \) by this sequence is
\[
(H \cdot M)_n = M_0 + \sum_{j=1}^{n} 1_{\{T \geq j\}} \Delta M_j,
\]
\[
= M_0 + \sum_{j=1}^{T \wedge n} \Delta M_j = M_{T \wedge n},
\]
as a consequence, if \( \{M_n\} \) is a (sub)martingale, the stopped process \( \{M_n^T\} := \{M_{T \wedge n}\} \) will be a (sub)martingale.

**Theorem 2.4.1 (Optional Stopping Theorem)** Suppose that \( \{M_n\} \) is a submartingale and \( S \leq T < m \) are two stopping times bounded by a fixed time \( m \).
\[
E(M_T | \mathcal{F}_S) \geq M_S
\]
with equality in the martingale case.

**Proof.** We make the proof only in the martingale case. Notice first that \( M_T \) is integrable because
\[
|M_T| \leq \sum_{n=0}^{m} |M_n|.
\]
Consider the predictable process \( H_n = 1_{\{S < n \leq T\} \cap A} \), where \( A \in \mathcal{F}_S \). Notice that \( \{H_n\} \) is predictable because
\[
\{S < n \leq T\} \cap A = \{T < n\}^c \cap (\{S \leq n - 1\} \cap A) \in \mathcal{F}_{n-1}.
\]
Moreover, the random variables $H_n$ are nonnegative and bounded by one. Therefore by Proposition 2.2.2, $(H \cdot M)_n$ is a martingale. We have
\[
(H \cdot M)_0 = M_0 \\
(H \cdot M)_n = M_0 + 1_A(M_T - M_S).
\]
The martingale property of $(H \cdot M)_n$ implies that $E(1_A(M_T - M_S)) = 0$ for all $A \in \mathcal{F}_S$ and this implies that $E(M_T|\mathcal{F}_S) = M_S$, because $M_S$ is $\mathcal{F}_S$-measurable.

**Theorem 2.4.2 (Doob’s Maximal Inequality)** Suppose that $\{M_n\}$ is a submartingale and $\lambda > 0$. Then
\[
P(\sup_{0 \leq n \leq N} M_n \geq \lambda) \leq \frac{1}{\lambda} E(M_N 1_{\{\sup_{0 \leq n \leq N} M_n \geq \lambda\}}) \leq \frac{1}{\lambda} E(M_N^+)\]

**Proof.** Consider the stopping time
\[
T = \inf\{n \geq 0, M_n \geq \lambda\} \wedge N.
\]
Then, by the Optional Stopping Theorem,
\[
E(M_N) \geq E(M_T) = E(M_T 1_{\{\sup_{0 \leq n \leq N} M_n \geq \lambda\}}) \\
+ E(M_T 1_{\{\sup_{0 \leq n \leq N} M_n < \lambda\}}) \\
\geq E(\lambda 1_{\{\sup_{0 \leq n \leq N} M_n \geq \lambda\}}) + E(M_N 1_{\{\sup_{0 \leq n \leq N} M_n < \lambda\}}).
\]

**Corollary 2.4.1** If $\{M_n\}$ is a martingale and $p \geq 1$, and $E(|M_n|^p) < \infty$ then
\[
P(\sup_{0 \leq n \leq N} M_n \geq \lambda) \leq \frac{1}{\lambda^p} E(|M_N|^p).
\]

**Remark 2.4.1** Note that the last inequality is a generalization of the Chebyshev inequality.

### 2.5 The Snell envelope and optimal stopping

Let $(Y_n)$ be an adapted process (to $(\mathcal{F}_n)$) with finite expectation, define
\[
X_N = Y_N \\
X_n = \max(Y_n, E(X_{n+1}|\mathcal{F}_n)), \quad 0 \leq n \leq N - 1,
\]
we say that $(X_n)$ is the Snell envelope of $(Y_n)$.

**Proposition 2.5.1** The sequence $(X_n)$ is the smallest a supermartingale that dominates the sequence $(Y_n)$. 
2.5. THE SNELL ENVELOPE AND OPTIMAL STOPPING

Proof. \((X_n)\) is adapted and by construction
\[ E(X_{n+1}|\mathcal{F}_n) \leq X_n. \]
Let \((T_n)\) be another supermartingale that dominates \((Y_n)\), then \(T_N \geq Y_N = X_N\).
Assume that \(T_{n+1} \geq X_{n+1}\). Then, by the monotony of the expectation and since \((T_n)\) is a supermartingale
\[ T_n \geq E(T_{n+1}|\mathcal{F}_n) \geq E(X_{n+1}|\mathcal{F}_n) \]
moreover \((T_n)\) dominates \((Y_n)\), so
\[ T_n \geq \max(Y_n, E(X_{n+1}|\mathcal{F}_n)) = X_n \]

Remark 2.5.1 Fixed \(\omega \) if \(X_n\) is strictly greater than \(Y_n\), \(X_n = E(X_{n+1}|\mathcal{F}_n)\) so \(X_n\) behaves, until this \(n\) as a martingale, this indicates that if we "stop" \(X_n\) properly we can have a martingale.

Proposition 2.5.2 The random variable
\[ \nu = \inf\{n \geq 0, X_n = Y_n\} \]
is a stopping time and \((X_\nu)\) is a martingale.

Proof.
\[ \{\nu = n\} = \{X_0 > Y_0\} \cap ... \cap \{X_{n-1} > Y_{n-1}\} \cap \{X_n = Y_n\} \in \mathcal{F}_n. \]
And
\[ X_\nu^n = X_0 + \sum_{j=1}^n 1_{\{j \leq \nu\}}(X_j - X_{j-1}) \]
therefore
\[ X_{\nu+1}^\nu - X_\nu^n = 1_{\{n+1 \leq \nu\}}(X_{n+1} - X_n) \]
and
\[ E(X_{\nu+1}^\nu - X_\nu^n|\mathcal{F}_n) = 1_{\{n+1 \leq \nu\}}E(X_{n+1} - X_n|\mathcal{F}_n) \]
\[ = \begin{cases} 0 & \text{if } \nu \leq n \text{ since the indicator vanishes} \\ 0 & \text{if } \nu > n \text{ since in such a case } X_n = E(X_{n+1}|\mathcal{F}_n) \end{cases} \]

We denote \(\tau_{n,N}\) stopping times with values in \(\{n, n+1, ..., N\}\).

Corollary 2.5.1
\[ X_0 = E(Y_\nu|\mathcal{F}_0) = \sup_{\tau \in \tau_{0,N}} E(Y_\tau|\mathcal{F}_0) \]
CHAPTER 2. DISCRETE TIME MARTINGALES

Proof. \((X_\nu^n)\) is a martingale and consequently

\[
X_0 = E(X_\nu^n|\mathcal{F}_0) = E(X_{N\wedge \nu}|\mathcal{F}_0)
= E(X_\nu|\mathcal{F}_0) = E(Y_\nu|\mathcal{F}_0).
\]

On the other hand \((X_n)\) is supermartingale and then \((X_\tau^n)\) as well for all \(\tau \in \tau_{0,N}\), so

\[
X_0 \geq E(X_N^\tau|\mathcal{F}_0) = E(X_\tau|\mathcal{F}_0) \geq E(Y_\tau|\mathcal{F}_0),
\]

therefore

\[
E(Y_\nu|\mathcal{F}_0) \geq E(Y_\tau|\mathcal{F}_0), \quad \forall \tau \in \tau_{0,N}
\]

Remark 2.5.2 Analogously we could prove

\[
X_n = E(Y_{\nu_n}|\mathcal{F}_n) = \sup_{\tau \in \tau_{n,N}} E(Y_\tau|\mathcal{F}_n),
\]

where

\[
\nu_n = \inf\{j \geq n, X_j = Y_j\}
\]

Definition 2.5.1 A stopping time \(\nu\) is said to be optimal for the sequence \((Y_n)\) if

\[
E(Y_\nu|\mathcal{F}_0) = \sup_{\tau \in \tau_{0,N}} E(Y_\tau|\mathcal{F}_0).
\]

Remark 2.5.3 The stopping time \(\nu = \inf\{n, X_n = Y_n\}\) (where \(X\) is the Snell envelope of \(Y\)) is then an optimal stopping time for \(Y\). We shall see the it is the smallest optimal stopping time.

The following theorem characterize the optimal stopping times.

Theorem 2.5.1 \(\tau\) is an optimal stopping time if and only if

\[
\begin{cases}
X_\tau = Y_\tau \\
(X_\tau^n) \text{ is a martingale}
\end{cases}
\]

Proof. If \((X_\tau^n)\) is a martingale and \(X_\tau = Y_\tau\)

\[
X_0 = E(X_N^n|\mathcal{F}_0) = E(X_{N\wedge \tau}|\mathcal{F}_0)
= E(X_\tau|\mathcal{F}_0) = E(Y_\tau|\mathcal{F}_0).
\]

On the other hand for all stopping time \(\pi\), \((X_\pi^n)\) is a supermartingale, so

\[
X_0 \geq E(X_N^\tau|\mathcal{F}_0) = E(X_\tau|\mathcal{F}_0) \geq E(Y_\tau|\mathcal{F}_0).
\]

Reciprocally, we know, by the previous corollary, that \(X_0 = \sup_{\tau \in \tau_{0,N}} E(Y_\nu|\mathcal{F}_0)\).

Then, if \(\tau\) is optimal

\[
X_0 = E(Y_\tau|\mathcal{F}_0) \leq E(X_\tau|\mathcal{F}_0) \leq X_0,
\]
2.6 Decomposition of Supermartingales

where the last inequality is due to the fact that \((X^\tau_n)\) is a supermartingale. So, we have

\[ E(X^\tau - Y^\tau | \mathcal{F}_n) = 0 \]

and since \(X^\tau - Y^\tau \geq 0\), we conclude that \(X^\tau = Y^\tau\).

Now we can also see that \((X^\tau_n)\) is a martingale. We know that it is a supermartingale, then

\[ X_0 \geq E(X^\tau_n | \mathcal{F}_0) \geq E(X^\tau_N | \mathcal{F}_0) = E(X^\tau | \mathcal{F}_0) = X_0 \]

as we saw before. Then, for all \(n\)

\[ E(X^\tau_n - E(X^\tau_n | \mathcal{F}_n) | \mathcal{F}_n) = 0, \]

and since \((X^\tau_n)\) is supermartingale,

\[ X^\tau_n \geq E(X^\tau_n | \mathcal{F}_n) = E(X^\tau | \mathcal{F}_n) \]

therefore \(X^\tau_n = E(X^\tau | \mathcal{F}_n)\). ■

2.6 Decomposition of Supermartingales

**Proposition 2.6.1** Any supermartingale \((X_n)\) has a unique decomposition:

\[ X_n = M_n - A_n \]

where \((M_n)\) is a martingale and \((A_n)\) is non-decreasing predictable with \(A_0 = 0\).

**Proof.** It is enough to write

\[ X_n = \sum_{j=1}^{n} (X_j - E(X_j | \mathcal{F}_{j-1})) - \sum_{j=1}^{n} (X_{j-1} - E(X_j | \mathcal{F}_{j-1})) + X_0 \]

and to identify

\[ M_n = \sum_{j=1}^{n} (X_j - E(X_j | \mathcal{F}_{j-1})) + X_0, \]

\[ A_n = \sum_{j=1}^{n} (X_{j-1} - E(X_j | \mathcal{F}_{j-1})) \]

where we define \(M_0 = X_0\) and \(A_0 = 0\). So \((M_n)\) is a martingale:

\[ M_n - M_{n-1} = X_n - E(X_n | \mathcal{F}_{n-1}), \quad 1 \leq n \leq N \]

in such a way that

\[ E(M_n - M_{n-1} | \mathcal{F}_{n-1}) = 0, \quad 1 \leq n \leq N. \]
Finally since \((X_n)\) is supermartingale
\[
A_n - A_{n-1} = X_{n-1} - E(X_n | \mathcal{F}_{n-1}) \geq 0, \quad 1 \leq n \leq N.
\]

Now we can see the uniqueness. If
\[
M_n - A_n = M'_n - A'_n, \quad 0 \leq n \leq N
\]
we have
\[
M_n - M'_n = A_n - A'_n, \quad 0 \leq n \leq N,
\]
but then since \((M_n)\) and \((M'_n)\) are martingales and \((A_n)\) and \((A'_n)\) predictable, it turns out that
\[
A_{n-1} - A'_{n-1} = M_{n-1} - M'_{n-1} = E(M_n - M'_n | \mathcal{F}_{n-1})
\]
\[
= E(A_n - A'_n | \mathcal{F}_{n-1}) = A_n - A'_n, \quad 1 \leq n \leq N,
\]
that is
\[
A_N - A'_N = A_{N-1} - A'_{N-1} = ... = A_0 - A'_0 = 0,
\]
since by hypothesis \(A_0 = A'_0 = 0.\) ■

This decomposition is known as the Doob decomposition.

**Proposition 2.6.2** The biggest optimal stopping time for \((Y_n)\) is given by
\[
\nu_{\text{max}} = \begin{cases} 
N & \text{si } A_N = 0 \\
\inf\{n, A_{n+1} > 0\} & \text{si } A_N > 0
\end{cases}
\]
where \((X_n),\) Snell envelope of \((Y_n),\) has a Doob decomposition \(X_n = M_n - A_n.\)

**Proof.** \(\{\nu_{\text{max}} = n\} = \{A_1 = 0, A_2 = 0, ..., A_n = 0, A_{n+1} > 0\} \in \mathcal{F}_n,\)
\(0 \leq n \leq N - 1, \{\nu_{\text{max}} = N\} = \{A_N = 0\} \in \mathcal{F}_{N-1}.\) So, it is a stopping time.

\[
X_{\nu_{\text{max}}} = X_{\cap \nu_{\text{max}}} = M_{\cap \nu_{\text{max}}} - A_{\cap \nu_{\text{max}}} = M_{\cap \nu_{\text{max}}}
\]
since \(A_{\cap \nu_{\text{max}}} = 0.\) Therefore \((X_{\nu_{\text{max}}})\) is a martingale. So, to see that this stopping time is optimal we have to prove that
\[
X_{\nu_{\text{max}}} = Y_{\nu_{\text{max}}}
\]

\[
X_{\nu_{\text{max}}} = \sum_{j=1}^{N-1} 1_{\{\nu_{\text{max}} = j\}} X_j + 1_{\{\nu_{\text{max}} = N\}} X_N
\]
\[
= \sum_{j=1}^{N-1} 1_{\{\nu_{\text{max}} = j\}} \max(Y_j, E(X_{j+1} | \mathcal{F}_j)) + 1_{\{\nu_{\text{max}} = N\}} Y_N,
\]
but in \(\{\nu_{\text{max}} = j\}, A_j = 0, A_{j+1} > 0\) so
\[
E(X_{j+1} | \mathcal{F}_j) = E(M_{j+1} | \mathcal{F}_j) - A_{j+1} < E(M_{j+1} | \mathcal{F}_j) = M_j = X_j
\]
therefore $X_j = Y_j$ en $\{\nu_{\text{max}} = j\}$ and consequently $X_{\nu_{\text{max}}} = Y_{\nu_{\text{max}}}$. Finally we see that is the biggest optimal stopping time. Let $\tau \geq \nu_{\text{max}}$ and $P\{\tau > \nu_{\text{max}}\} > 0$. Then
\[
E(X_\tau) = E(M_\tau) - E(A_\tau) = E(M_0) - E(A_\tau)
\]
\[
= X_0 - E(A_\tau) < X_0
\]
so $(X_{\tau \wedge n})$ cannot be a martingale. ❑

2.7 Martingale convergence theorems

Let $\{M_n\}_{n \geq 0}$ be a supermartingale and $a < b \in \mathbb{R}$, let $U_N[a, b](\omega)$ be the number of upcrossings made by $(M_n(\omega))$ by time $N$. Then define a predictable strategy $H_1 = 1_{(M_0 < a)}$ and $H_n = 1_{(M_{n-1} < a, H_{n-1} = 0)} + 1_{(M_{n-1} \leq b, H_{n-1} = 1)}$, for $n \geq 2$. Then
\[
(H \cdot M)_n := \sum_{j=1}^{n} H_j (M_j - M_{j-1})
\]
is a supermartingale that represents the gain by time $n$ in a game where the profit at time $n$ if the bet is 1 euro. It is easy to see (by a simple picture) that
\[
(H \cdot M)_n \geq (b - a)U_n[a, b] - (M_n - a)^-, n \geq 1
\]
then we have:

**Lemma 2.7.1** Let $\{M_n\}_{n \geq 0}$ be a supermartingale let $U_N[a, b](\omega)$ be the number of upcrossings by time $N$. Then
\[
(b - a)E(U_N[a, b]) \leq E((M_N - a)^-).
\]

**Proof.** The process $(H \cdot M)_n$ above is a supermartingale then $E((H \cdot M)_N) \leq 0$. ❑

**Corollary 2.7.1** Let $\{M_n\}_{n \geq 0}$ be a supermartingale such that $\sup_n E(|M_n|) < \infty$. Let $a < b \in \mathbb{R}$ and $U_\infty[a, b] := \lim_{n \to \infty} U_n[a, b]$, then
\[
(b - a)E(U_\infty[a, b]) \leq |a| + \sup_n E(|M_n|) < \infty,
\]
so that
\[
P(U_\infty[a, b] = \infty) = 0.
\]

**Proof.** By the previous lemma
\[
(b - a)E(U_N[a, b]) \leq |a| + E(|M_N|)
\]
and taking the limit when $N \to \infty$ and using the monotone convergence theorem we obtain the result. ❑
Theorem 2.7.1 Let \( \{ M_n \}_{n \geq 0} \) be a supermartingale such that \( \sup_n E(|M_n|) < \infty \). Then almost surely \( M_\infty := \lim_{n \to \infty} M_n \) exists and is finite.

**Proof.** Let \( \Lambda = \{ \omega, M_n(\omega) \) does not converge to \( a \) in \( [−\infty, +\infty] \} \)

\[
\Lambda = \cup_{a < b \in \mathbb{Q}} \{ \omega, \liminf M_n(\omega) < a < b < \limsup M_n(\omega) \}
\]

but \( \Lambda_{ab} \subset \{ U_\infty[a, b) = \infty \} \),

and by the previous corollary \( P(\Lambda_{ab}) = 0 \) and then \( P(\Lambda) = 0 \). So \( M_\infty \) exists in \( [−\infty, +\infty] \) a.s., but by Fatou’s lemma

\[
E(|M_\infty|) = E(\liminf |M_n|) \leq \liminf E(|M_n|) \leq \sup_n E(|M_n|) < \infty,
\]

so \( M_\infty \) is finite a.s. ■

**Corollary 2.7.2** If \( \{ M_n \}_{n \geq 0} \) is a non negative supermartingale then \( \lim_{n \to \infty} M_n \) exists and is finite almost surely.

**Proof.** \( E(|M_n|) = E(M_n) \leq E(M_0) \). ■

**Remark 2.7.1** It is important to note that it need no be true that \( M_n \to M_\infty \) in \( L^1 \).

**Example 2.7.1** Suppose that \( \{ \xi_n, n \geq 1 \} \) are i.i.d. centered random variables with distribution \( N(0, 2\sigma^2) \). Set \( M_0 = 1 \) and

\[
M_n = \exp \left( \sum_{j=1}^{n} \xi_j - n\sigma^2 \right).
\]

Then, \( \{ M_n \} \) is a nonnegative martingale such that \( M_n \to 0 \) a.s., by the strong law of large numbers, but \( E(M_n) = 1 \) for all \( n \).

**Theorem 2.7.2** Let \( \{ M_n \}_{n \geq 0} \) be a martingale, such that \( \sup_n E(M_n^2) < \infty \), then \( M_n \to M_\infty \) a.s. and in \( L^2 \).

**Proof.** Obviously \( \sup_n E(|M_n|) < \infty \) so we have almost sure convergence to \( M_\infty \), on the other hand by the orthogonality of the increments

\[
E(M_n^2) = E(M_0^2) + \sum_{k=1}^{n} E((M_k - M_{k-1})^2),
\]

so

\[
\sum_{k=1}^{\infty} E((M_k - M_{k-1})^2) < \infty,
\]
and applying Fatou’s lemma in

\[ E((M_{n+r} - M_n)^2) = \sum_{k=n+1}^{n+r} E((M_k - M_{k-1})^2), \]

when \( r \to \infty \), we have

\[ E((M_\infty - M_n)^2) \leq \sum_{k=n+1}^{\infty} E((M_k - M_{k-1})^2), \]

so

\[ \lim_{n \to \infty} E((M_\infty - M_n)^2) = 0. \]
Chapter 3

Markov Processes

3.1 Discrete-time Markov chains

**Definition 3.1.1** We say that \((X_n)_{n \geq 0}\) is a Markov chain with initial distribution \(\lambda := (\lambda_i, i \in I)\) and transition matrix \(P := (p_{ij}, i, j \in I)\)

(i) \(P\{X_0 = i\} = \lambda_i, i \in I\)
(ii) for all \(n \geq 0\), \(P(X_{n+1} = i_{n+1} | X_0 = i_0, X_1 = i_1, ..., X_n = i_n) = p_{i_n i_{n+1}}\).

Where \(I\) is a countable set. We say that \((X_n)_{n \geq 0}\) is Markov(\(\lambda, P\)) for short.

We regard \(\lambda\) as a row vector whose components are indexed by \(I\) and \(P\) a matrix whose entries are indexed by \(I \times I\). We write \(p_{ij}^{(n)} = (P^n)_{ij}\) for the \((i, j)\) entry in \(P^n\). We agree that \(P^0\) is the identity matrix, that is \((P^0)_{ij} = \delta_{ij}\).

**Proposition 3.1.1** Let \((X_n)_{n \geq 0}\) be Markov(\(\lambda, P\)). Then for all \(n, m \geq 0\)

\[
P(X_{n+m} = j | X_0 = i_0, ..., X_n = i) = P(X_{n+m} = j | X_n = i) = P(X_m = j | X_0 = i).
\]
The most general two-state chain has transition matrix

Example 3.1.1

Proof.

\[ P(X_{n+m} = j | X_n = i) = \frac{P(X_{n+m} = j, X_n = i)}{P(X_n = i)} = \frac{\sum_{i_0 \in I} \sum_{i_1 \in I} \cdots \sum_{i_{n+m-1} \in I} P(X_0 = i_0, \ldots, X_n = i)p_{i_{n+1} \cdots p_{i_{n+m-1}j}}}{\sum_{i_0 \in I} \sum_{i_1 \in I} \cdots \sum_{i_n \in I} P(X_0 = i_0, \ldots, X_n = i)} = \frac{\sum_{i_{n+1} \in I} \cdots \sum_{i_{n+m-1} \in I} \prod_{i_{n+1} \in I} \cdots \sum_{i_{n+m-1} \in I} P(X_0 = i, X_1 = i_{n+1}, \ldots, X_{m-1} = i_{n+m-1}, X_m = j)}{P(X_0 = i)} = \frac{P(X_0 = i, X_m = j)}{P(X_0 = i)} = P(X_m = j | X_0 = i) \]

\[ \square \]

Remark 3.1.1 Note also that

\[ P(X_{n+1} = j_1, X_{n+2} = j_2, \ldots, X_{n+m} = j_m | X_0 = i_0, \ldots, X_n = i) = P(X_1 = j_1, X_2 = j_2, \ldots, X_m = j_m | X_0 = i). \]

Proposition 3.1.2 Let \((X_n)_{n \geq 0}\) be Markov(\(\lambda, P\)). Then, for all \(n, m \geq 0\),

(i) \[ P(X_n = j) = (\lambda P^n)_j, \quad j \in I \]

(ii) \[ P_i(X_n = j) := P(X_n = j | X_0 = i) = P(X_{n+m} = j | X_m = i) = \left( p^{(n)} \right)_{ij}. \]

Proof. (i):

\[ P(X_n = j) = \sum_{i_0 \in I} \sum_{i_1 \in I} \cdots \sum_{i_{n-1} \in I} P(X_0 = i_0, \ldots, X_n = j) \]

\[ = \sum_{i_0 \in I} \sum_{i_1 \in I} \cdots \sum_{i_{n-1} \in I} P(X_n = j | X_{n-1} = i_{n-1}, \ldots, X_0 = i_0)P(X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \]

\[ = \sum_{i_0 \in I} \sum_{i_1 \in I} \cdots \sum_{i_{n-1} \in I} p_{i_{n-1}j}P(X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \]

\[ = \sum_{i_0 \in I} \sum_{i_1 \in I} \cdots \sum_{i_{n-1} \in I} P(X_n = j | X_{n-2} = i_{n-2}, \ldots, X_0 = i_0) \]

\[ = \cdots = \sum_{i_0 \in I} \sum_{i_1 \in I} \cdots \sum_{i_{n-1} \in I} \lambda_{i_0} p_{i_0i_1} \cdots p_{i_{n-2}i_{n-1}} p_{i_{n-1}j} = (\lambda P^n)_j. \]

(ii): \[ P(X_n = j | X_0 = i) = \sum_{i_1 \in I} \cdots \sum_{i_{n-1} \in I} p_{i_1i_2} \cdots p_{i_{n-1}j} = (P^n)_{ij} = \left( p^{(n)} \right)_{ij}. \]

Example 3.1.1 The most general two-state chain has transition matrix

\[ P = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix}. \]
3.1. DISCRETE-TIME MARKOV CHAINS

If we want to calculate $p_{11}^{(n)}$ we can use the relation $P^{n+1} = P^n P$, then

$$p_{11}^{(n+1)} = p_{11}^{(n)} (1 - \alpha) + p_{12}^{(n)} \beta.$$ 

We know that

$$p_{11}^{(n)} + p_{12}^{(n)} = P_1(X_n = 1 \text{ or } 2) = 1,$$

so

$$p_{11}^{(n+1)} = p_{11}^{(n)} (1 - \alpha - \beta) + \beta, n \geq 0$$

with $p_{11}^{(0)} = 1$. This has a unique solution

$$p_{11}^{(n)} = \left\{ \begin{array}{ll} \frac{\beta}{\alpha + \beta} + \frac{\alpha}{\alpha + \beta} (1 - \alpha - \beta)^n & \text{for } \alpha + \beta > 0 \\ 1 & \text{for } \alpha + \beta = 0. \end{array} \right.$$ 

3.1.1 Class structure

It is sometimes possible to split a Markov chain into smaller pieces, each of which is relatively easy to analyze, and which together we can understand the whole. This is done by identifying the communicating classes.

We say that $i$ leads to $j$ and write $i \rightarrow j$ if

$$P_i(X_n = j \text{ for some } n \geq 0) > 0.$$ 

We say that $i$ communicates with $j$ and write $i \leftrightarrow j$ if both $i \rightarrow j$ and $j \rightarrow i$.

**Proposition 3.1.3** The following are equivalent

(i) $i \rightarrow j$;

(ii) $p_{i_0 i_1} \cdots p_{i_{n-1} i_n} > 0$ for some states $i_0, i_1, ..., i_n$, with $i_0 = i$ and $i_n = j$;

(iii) $p_{ij}^{(n)} > 0$ for some $n \geq 0$.

**Proof.** (i)$\sim$(iii):

$$p_{ij}^{(n)} \leq P_i(X_n = j \text{ for some } n \geq 0) \leq \sum_{n=0}^{\infty} p_{ij}^{(n)}.$$ 

(ii)$\sim$ (iii):

$$p_{ij}^{(n)} = \sum_{i_1, i_2, ..., i_{n-1}} p_{i_1 i_2} \cdots p_{i_{n-1} j}.$$ 

It is clear that $i \leftrightarrow j$ is an equivalence relation. So, it partitions $I$ (the state space) into communicating classes. We say that a class $C$ is **closed** if

$$i \in C, i \rightarrow j \text{ imply } j \in C.$$ 

We cannot escape from a closed class. A state $i$ is **absorbing** if $\{i\}$ is a closed class. A chain or matrix $P$ is called **irreducible** if $I$ is a single class.
Example 3.1.2 Consider the stochastic matrix
\[
P = \begin{pmatrix}
\frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
\frac{1}{3} & 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 \\
0 & 0 & 0 & \frac{2}{3} & \frac{1}{2} & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 0
\end{pmatrix},
\]
the communicating classes associated with it are: \{1, 2, 3\}, \{4\} and \{5, 6\}, with only \{5, 6\} being closed.

3.1.2 Hitting times and absorption probabilities
Let \(\tau^A\) be the hitting time of a subset \(A \subseteq I\):
\[
\tau^A = \inf\{n \geq 0, X_n \in A\}.
\]
One purpose of this subsection is to learn how to calculate the probability, starting from \(i\), that \((X_n)_{n \geq 0}\) ever hits \(A\) that is
\[
h_i^A := P_i(\tau^A < \infty).
\]
When \(A\) is a closed class this probability is named absorption probability. The other purpose is to calculate the mean time taken for \((X_n)_{n \geq 0}\) to reach \(A\):
\[
k_i^A := E_i(\tau^A) = \sum_{1 \leq n < \infty}nP_i(\tau^A = n) + \infty P_i(\tau^A = \infty).
\]

Proposition 3.1.4
\[
E_i(\tau^A) = \sum_{1 \leq n \leq \infty} P_i(\tau^A \geq n)
\]

Proof.
\[
P_i(\tau^A \geq n) = \sum_{n \leq m \leq \infty} P_i(\tau^A = m),
\]
then
\[
\sum_{1 \leq n \leq \infty} P_i(\tau^A \geq n) = \sum_{1 \leq n \leq \infty} \sum_{n \leq m \leq \infty} P_i(\tau^A = m) = \sum_{1 \leq m \leq \infty} \sum_{1 \leq n \leq m} P_i(\tau^A = m) = \sum_{1 \leq m \leq \infty} mP_i(\tau^A = m).
\]
Example 3.1.3 Consider the following transition matrix

\[
P = \begin{pmatrix}
1 & 0 & 0 & 0 \\
\frac{1}{2} & 0 & \frac{1}{2} & 0 \\
0 & \frac{1}{2} & 0 & \frac{1}{2} \\
0 & 0 & 0 & 1 \\
\end{pmatrix},
\]

Introduce \( h_i := h_i^{[4]} \) and \( k_i := k_i^{[1,4]} \). Then it is clear that \( h_1 = 0, h_4 = 1 \) and \( k_1 = k_4 = 0 \). Also we have

\[
h_2 = \frac{1}{2} h_1 + \frac{1}{2} h_3, \quad k_2 = 1 + \frac{1}{2} k_1 + \frac{1}{2} k_3,
\]

and

\[
h_3 = \frac{1}{2} h_2 + \frac{1}{2} h_4, \quad k_3 = 1 + \frac{1}{2} k_2 + \frac{1}{2} k_4.
\]

Therefore

\[
h_2 = \frac{1}{2} h_3, \quad k_2 = 1 + \frac{1}{2} k_3
\]

\[
h_3 = \frac{1}{2} h_2 + \frac{1}{2}, \quad k_3 = 1 + \frac{1}{2} k_2
\]

and then

\[
h_2 = \frac{1}{3}, \quad h_3 = \frac{2}{3}
\]

\[
k_2 = k_3 = 2.
\]

Proposition 3.1.5 The vector of hitting probabilities \( h^A = (h_i^A : i \in I) \) is the minimal non-negative solution to the system of linear equations

\[
\begin{aligned}
h_i^A &= 1 & \text{for } i \in A \\
h_i^A &= \sum_{j \in I} p_{ij} h_j^A & \text{for } i \notin A.
\end{aligned}
\]

where minimality means that if \( x = (x_i : i \in I) \) is another solution with \( x_i \geq 0 \) for all \( i \), then \( x_i \geq h_i^A \), for all \( i \).

Proof. First we prove that \( h^A \) satisfies (3.1). If \( X_0 = i \in A \) then \( \tau^A = 0 \), so \( h_i^A = 1 \). If \( X_0 = i \notin A \), then \( \tau^A \geq 1 \) and by the Markov property

\[
P_i(\tau^A < \infty | X_1 = j) = P_j(\tau^A < \infty) = h_j^A.
\]

Then

\[
h^A_i = \sum_{j \in I} P_i(\tau^A < \infty, X_1 = j)
\]

\[
= \sum_{j \in I} P_i(\tau^A < \infty | X_1 = j) P_i(X_1 = j)
\]

\[
= \sum_{j \in I} P_j(\tau^A < \infty) P_i(X_1 = j) = \sum_{j \in I} p_{ij} h_j^A.
\]
Suppose now that \( x = (x_i : i \in I) \) is another solution to \((3.1)\) with \( x_i \geq 0 \). Then \( h_k^A = x_i = 1 \) for \( i \in A \) consequently, if \( i \notin A \)

\[
x_i = \sum_{j \in I} p_{ij} x_j = \sum_{j \in A} p_{ij} + \sum_{j \notin A} p_{ij} x_j.
\]

Now if we substitute for \( x_j \) we obtain

\[
x_i = \sum_{j \in A} p_{ij} + \sum_{j \notin A} p_{ij} \left( \sum_{k \in A} p_{jk} + \sum_{k \notin A} p_{ik} x_k \right)
\]

\[
= \sum_{j \in A} p_{ij} + \sum_{j \notin A} p_{ij} \sum_{k \in A} p_{jk} + \sum_{j \notin A} \sum_{k \notin A} p_{ij} p_{ik} x_k
\]

\[
= P_i(X_1 \in A) + P_i(X_1 \notin A, X_2 \in A) + \sum_{j \notin A} \sum_{k \notin A} p_{ij} p_{ik} x_k.
\]

By repeating the substitution, after \( n \) steps we have

\[
x_i = P_i(X_1 \in A) + ... + P_i(X_1 \notin A, ..., X_{n-1} \notin A, X_n \in A) + \sum_{j_1 \notin A} ... \sum_{j_n \notin A} p_{i j_1} p_{j_1 j_2} \cdots p_{j_{n-1} j_n} x_{j_n}.
\]

So, since the last term is non-negative, for all \( n \geq 1 \)

\[
x_i \geq P_i(X_1 \in A) + ... + P_i(X_1 \notin A, ..., X_{n-1} \notin A, X_n \in A) = P_i(\tau^A \leq n).
\]

Finally

\[
x_i \geq \lim_{n \to \infty} P_i(\tau^A \leq n) = P_i(\tau^A < \infty) = h_i^A.
\]

\[\text{Proposition 3.1.6} \quad \text{The vector of mean hitting times } k^A = (k_i^A : i \in I) \text{ is the minimal non-negative solution to the system of linear equations}
\]

\[
\begin{cases}
  k_i^A = 0 & \text{for } i \in A \\
  k_i^A = 1 + \sum_{j \notin A} p_{ij} k_j^A & \text{for } i \notin A.
\end{cases}
\]

\[\text{Proof.} \quad \text{First we prove that } k^A \text{ satisfies the system } (3.2). \quad \text{If } i \in A \text{ then } \tau^A = 0 \text{ and } k_i^A = 0. \text{ If } X_0 = i \notin A, \text{ then } \tau^A \geq 1 \text{ and by the Markov property}
\]

\[
E_i(\tau^A | X_1 = j) = 1 + E_j(\tau^A).
\]

Then

\[
E_i(\tau^A) = \sum_{j \in I} E_i(\tau^A 1_{\{X_1 = j\}}) = \sum_{j \in I} E_i(\tau^A | X_1 = j) P_i(X_1 = j)
\]

\[
= 1 + \sum_{j \in I} E_j(\tau^A) p_{ij} = 1 + \sum_{j \notin A} p_{ij} k_j^A.
\]
Now, assume that \( y = (y_i, i \in \mathcal{I}) \) is a solution of the system (3.2). Then

\[
y_i = 1 + \sum_{j \notin A} p_{ij} y_j = 1 + \sum_{j \notin A} p_{ij} (1 + \sum_{k \notin A} p_{jk} y_k)
\]

\[
= 1 + \sum_{j \notin A} p_{ij} + \sum_{j \notin A} \sum_{k \notin A} p_{ij} p_{jk} y_k
\]

\[
= P_i(\tau^A \geq 1) + P_i(\tau^A \geq 2) + \sum_{j \notin A} \sum_{k \notin A} p_{ij} p_{jk} y_k
\]

\[
= \ldots = P_i(\tau^A \geq 1) + \ldots + P_i(\tau^A \geq n)
\]

\[
+ \sum_{j \notin A} \ldots \sum_{j_n \notin A} p_{ij_1} \ldots p_{j_{n-1} j_n} y_{j_n},
\]

and

\[
y_i \geq \sum_{1 \leq n \leq \infty} P_i(\tau^A \geq n).
\]

\[\text{(Gambler’s ruin)}\]

Consider the transition probabilities

\[
p_{00} = 1 \quad p_{i, i-1} = q > 0, \quad p_{i, i+1} = p = 1 - q > 0, \quad \text{for } i = 1, 2, \ldots.
\]

Imagine you enter in the casino with 1 euro and you can win 1 euro with probability \( p \) or to lose it with probability \( q \). The resources of the casino are regarded as infinite, so there is no upper bound to your fortune. But which is the probability you finish broke?

Set \( h_i = P_i(\tau^{\{0\}} < \infty) \), then \( h_i \) is the minimal solution of the system

\[
h_0 = 1,
\]

\[
h_i = qh_{i-1} + ph_{i+1}, \quad i = 1, 2, \ldots
\]

If \( p \neq q \) this recurrence equation has a general solution

\[
h_i = A + B \left( \frac{q}{p} \right)^i.
\]

If \( p < q \) (quite often) the restriction \( 0 \leq h_i \leq 1 \) forces \( B \) to be zero so \( h_i = 1 \) for all \( i \). If \( p > q \), since \( h_0 = 1 \), we have a family of solutions

\[
h_i = \left( \frac{q}{p} \right)^i + A (1 - \left( \frac{q}{p} \right)^i),
\]

and since \( h_i \geq 0 \) we must have \( A \geq 0 \), then the minimal solution is \( h_i = \left( \frac{q}{p} \right)^i \).

Finally if \( p = q \) the recurrent relation has a general solution

\[
h_i = A + Bi
\]
and again the restriction \( 0 \leq h_i \leq 1 \) forces \( B \) to be zero, so \( h_i = 1 \) for all \( i \). Thus, even if you find a fair casino and you have a lot of money, you are certain to end up broke.

**Proposition 3.1.7 (Strong Markov property)** Let \( (X_n)_{n \geq 0} \) be Markov(\( \lambda, P \)). And \( \tau \) a finite stopping time associated to \( (X_n)_{n \geq 0} \) and \( A \in \mathcal{F}_\tau \) then for all \( m \geq 0 \)

\[
P(X_{\tau+m} = j|A, X_\tau = i) = P(X_m = j|X_0 = i).
\]

That is \( (X_{\tau+n})_{n \geq 0} \) is Markov(\( \mu, P \)) where \( \mu_i = P(X_\tau = i) \).

**Proof.**

\[
P(X_{\tau+m} = j|A, X_\tau = i) = \frac{P(X_{\tau+m} = j, A, X_\tau = i)}{P(A, X_\tau = i)} = \frac{\sum_{n < \infty} P(X_{\tau+m} = j, A, X_\tau = i, \tau = n)}{\sum_{n < \infty} P(A, \tau = i, \tau = n)} = \frac{\sum_{n < \infty} P(X_{n+m} = j, A, X_n = i, \tau = n)}{\sum_{n < \infty} P(A, X_n = i, \tau = n)} = \frac{\sum_{n < \infty} P(X_m = j|X_0 = i)P(A, X_n = i, \tau = n)}{\sum_{n < \infty} P(A, X_n = i, \tau = n)} = \frac{P(X_m = j|X_0 = i)\sum_{n < \infty} P(A, X_n = i, \tau = n)}{\sum_{n < \infty} P(A, X_n = i, \tau = n)} = \frac{P(X_m = j|X_0 = i)}{\sum_{n < \infty} P(A, X_n = i, \tau = n)}
\]

**Remark 3.1.2** Note also that if \( \tau \) is a finite stopping time associated to \( (X_n)_{n \geq 0} \) and \( A \in \mathcal{F}_\tau \) then for all \( m \geq 0 \)

\[
P(X_{\tau+1} = j_1, X_{\tau+2} = j_2, ..., X_{\tau+m} = j_m|A, X_\tau = i)
\]

\[
= P(X_1 = j_1, X_2 = j_2, ..., X_m = j_m|X_0 = i).
\]

In particular assume that

\[
\tau_0 = \inf\{n \geq 0, X_n \in J \subset I\}
\]

and for \( m = 0, 1, 2, ... \)

\[
\tau_{m+1} = \inf\{n > \tau_m, X_n \in J \subset I\},
\]

then

\[
P(X_{\tau_{m+1}} = j|X_{\tau_0} = i_0, X_{\tau_1} = i_1, ..., X_{\tau_m} = i) = P(X_{\tau_1} = j|X_{\tau_0} = i),
\]

where for \( i, j \in J \)

\[
P(X_{\tau_1} = j|X_{\tau_0} = i) = P_i(X_n = j \text{ for some } n \geq 1)
\]
3.1.3 Recurrence and transience

Let \((X_n)_{n \geq 0}\) be a Markov chain with transition matrix \(P\). We say that a state \(i\) is recurrent if

\[ P_i(X_n = i, \text{ for infinitely many } n) = 1. \]

We say that \(i\) is transient if

\[ P_i(X_n = i, \text{ for infinitely many } n) = 0. \]

Recall that the first passage time to state \(i\) is the random variable \(\tau^{(i)}\) defined as

\[ \tau^{(i)} = \inf\{n \geq 1, X_n = i\}, \]

then we can define inductively the \(r\)th passage time \(\tau^{(i)}_r\) to state \(i\) for \(r = 1, 2, \ldots\) by

\[ \tau^{(i)}_1 = \tau^{(i)}, \quad \tau^{(i)}_{r+1} = \inf\{n \geq \tau^{(i)}_r + 1, X_n = i\}. \]

The length for the \(r\)th excursion, for \(r \geq 2\), is given by

\[ \delta^{(i)}_r := \begin{cases} \tau^{(i)}_r - \tau^{(i)}_{r-1} & \text{if } \tau^{(i)}_{r-1} < \infty \\ 0 & \text{if } \tau^{(i)}_{r-1} = \infty. \end{cases} \]

Lemma 3.1.1 For \(r \geq 2\), conditional on \(\tau^{(i)}_{r-1} < \infty\) and for all \(A \in \mathcal{F}_{\tau^{(i)}_{r-1}}\) we have

\[ P(\delta^{(i)}_r = n | \tau^{(i)}_{r-1} < \infty, A) = P_i(\tau^{(i)} = n). \]

Proof. By the strong Markov property and conditional on \(\{\tau^{(i)}_{r-1} < \infty\}\), \((X_{\tau^{(i)}_{r-1} + n})_{n \geq 0}\) is a Markov chain with transition matrix \(P = (p_{ij})\), \(p_{ij} = P(X_1 = j | X_0 = i)\) and \(\lambda_j = \delta_{ij}\) for all \(j \in I\). Then \(\delta^{(i)}_r\) is the first passage time of \((X_{\tau^{(i)}_{r-1} + n})_{n \geq 0}\) to state \(i\), so

\[ P(\delta^{(i)}_r = n | \tau^{(i)}_{r-1} < \infty, A) = P(\tau^{(i)} = n | X_0 = i). \]

Let us introduce the number of visits \(V_i\) to \(i\):

\[ V_i = \sum_{n=0}^{\infty} 1_{\{X_n = i\}}, \]

note that

\[ E_i(V_i) = \sum_{n=0}^{\infty} E_i(1_{\{X_n = i\}}) = \sum_{n=0}^{\infty} P_i(X_n = i) = \sum_{n=0}^{\infty} p_i^{(n)}. \]

Also we can compute the distribution of \(V_i\) under \(P_i\) in terms of the return probability

\[ f_i = P_i(\tau^{(i)} < \infty). \]
Lemma 3.1.2 For $r \geq 0$ we have $P_i(V_i > r) = f^r_i$ (we assume $0^0 = 1$).

Proof. For $P_i(V_i > 0) = 1$ since $X_0 = i$, so the result is true for $r = 0$. 

Assume $\{X_0 = i, V_i > r\} = \{X_0 = i, \tau^{(i)} < \infty\}$, assume the result is true for $r$, then

\[
P_i(V_i > r + 1) = P_i(\tau^{(i)} < \infty)
\]

\[
= P_i(\delta^{(i)}_{r+1} < \infty | \tau^{(i)} < \infty) P_i(\tau^{(i)} < \infty)
\]

\[
= P_i(\tau^{(i)} < \infty) f^r_i.
\]

Theorem 3.1.1 The following dichotomy holds:

(i) if $P_i(\tau^{(i)} < \infty) = 1$ then $i$ is recurrent and $\sum_{n=0}^{\infty} p^{(n)}_{ii} = \infty$

(ii) if $P_i(\tau^{(i)} < \infty) < 1$ then $i$ is transient and $\sum_{n=0}^{\infty} p^{(n)}_{ii} < \infty$.

Proof. 

\[
P_i(V_i = \infty) = \lim_{n \to \infty} P_i(V_i > n) = f^n_i,
\]

then if $f_i := P_i(\tau^{(i)} < \infty) = 1$, then $P_i(V_i = \infty) = 1$ and $i$ is recurrent and

\[
\sum_{n=0}^{\infty} p^{(n)}_{ii} = E_i(V_i) = \infty.
\]

On the other hand if $f_i < 1$

\[
\sum_{n=0}^{\infty} p^{(n)}_{ii} = E_i(V_i) = \sum_{n=0}^{\infty} P_i(V > n) = \sum_{n=0}^{\infty} f^n_i
\]

\[
= \frac{1}{1 - f_i} < \infty,
\]

so $P_i(V_i = \infty) = 0$ and $i$ is transient. Recurrence and trasience are class properties.

Theorem 3.1.2 Let $C$ a communicating class. Then either all states in $C$ are transient or recurrent.

Proof. Take a pair of states $i, j \in C$ and assume that $i$ is transient. Then there exist $n, m$ such that $p^{(n)}_{ij} > 0$ and $p^{(m)}_{ji} > 0$, and for all $r \geq 0$

\[
p^{(n+m+r)}_{ii} \geq p^{(n)}_{ij} p^{(r)}_{jj} p^{(m)}_{ji},
\]
3.1. DISCRETE-TIME MARKOV CHAINS

then
\[ \sum_{r=0}^{\infty} p_{jj}^{(r)} \leq \frac{1}{p_{ij}^{(n)} p_{ji}} \sum_{r=0}^{\infty} p_{ii}^{(n+m+r)} < \infty, \]
so by the previous Theorem \( j \) is also transient. 

**Theorem 3.1.3** Every recurrent class is closed.

**Proof.** Let \( C \) a class that is not closed. Then there exist \( i \in C, j \notin C \) and \( m \geq 1 \) such that
\[ P_i(X_m = j) > 0. \]
Since we have that
\[ P_i(X_m = j, X_n = i \text{ for infinitely many } n) = 0, \]
this implies that
\[ P_i(X_n = i \text{ for infinitely many } n) < 1, \]
so \( i \) is not recurrent, and so neither is \( C \). 

**Theorem 3.1.4** Every finite closed class is recurrent.

**Proof.** Suppose that \( C \) is close and that \( (X_n)_{n \geq 0} \) starts in \( C \). Then for some \( i \in C \)
\[ P(X_n = i \text{ for infinitely many } n) > 0, \]
since the class is finite. Let \( \tau = \inf\{n \geq 0, X_n = i\} \) then
\[ P(X_n = i \text{ for infinitely many } n) \]
\[ = P(\tau < \infty) P(X_{\tau + m} = i \text{ for infinitely many } m | \tau < \infty) \]
\[ = P(\tau < \infty) P_i(X_m = i \text{ for infinitely many } m) > 0. \]
So this shows that \( i \) is not transient. 

### 3.1.4 Recurrence and transient in a random walk

Consider the transition probabilities in \( \mathbb{Z} \)
\[ p_{i,i-1} = q > 0, \ p_{i,i+1} = p = 1 - q > 0, \text{ for } i \in \mathbb{Z}, \]
that is a simple random walk on \( \mathbb{Z} \). It is clear that
\[ p_{00}^{(2n+1)} = 0, \]
and it is easy to see that
\[ p_{00}^{(2n)} = \binom{2n}{n} p^n q^n. \]
But Stirling’s formula
\[ n! \sim \sqrt{2\pi nn^e}^{-n}, \quad \text{as } n \to \infty \]
so
\[ p_{00}^{(2n)} = \left(\frac{(2n)!}{(n)!^2}\right) (pq)^n \sim \frac{1}{\sqrt{n}} \frac{(4pq)^n}{\sqrt{n}}, \quad \text{as } n \to \infty. \]
If \( p = q = \frac{1}{2} \), \( p_{00}^{(2n)} \sim \frac{1}{\sqrt{n}} \frac{1}{\sqrt{n}}, \) and since
\[ \sum_{n=1}^{\infty} \frac{1}{\sqrt{n}} = \infty, \]
we have that
\[ \sum_{n=0}^{\infty} p_{00}^{(n)} = \infty, \]
so the random walk is recurrent. If, on the contrary, \( p \neq q \), then \( 4pq = r < 1 \) and \( p_{00}^{(2n)} \sim \frac{1}{\sqrt{n}} \frac{r^n}{\sqrt{n}}, \) now since
\[ \sum_{n=1}^{\infty} \frac{r^n}{\sqrt{n}} < \infty, \]
we have that
\[ \sum_{n=0}^{\infty} p_{00}^{(n)} < \infty, \]
and the random walk is transient.

### 3.1.5 Invariant distributions
Many of the long-time properties of Markov chains are connected with the notion of an invariant distribution or measure. We say that \( \lambda \) is invariant if
\[ \lambda P = \lambda. \]

**Theorem 3.1.5** Let \( (X_n)_{n \geq 0} \) be \( \text{Markov}(\lambda, P) \) and suppose that \( \lambda \) is invariant then \( (X_{m+n})_{n \geq 0} \) is also \( \text{Markov}(\lambda, P). \)

**Proof.** By the Markov property \( (X_{m+n})_{n \geq 0} \) has the same transition matrix with initial distribution \( P(X_m = i) = (\lambda P^m)_i = \lambda_i. \)

**Theorem 3.1.6** Let \( I \) be finite. Suppose for some \( i \in I \) that
\[ p_{ij}^{(n)} \to \pi_j \quad \text{as } n \to \infty \quad \text{for all } j \in I. \]
Then \( \pi = \{\pi_j, j \in I\} \) is an invariant distribution.
Proof. We have
\[ \sum_{j \in I} \pi_j = \sum_{j \in I} \lim_{n \to \infty} p_{ij}^{(n)} = \lim_{n \to \infty} \sum_{j \in I} p_{ij}^{(n)} = 1, \]
and
\[ \pi_j = \lim_{n \to \infty} p_{ij}^{(n)} = \lim_{n \to \infty} \sum_{k \in I} p_{ik}^{(n-1)} p_{kj} = \sum_{k \in I} \lim_{n \to \infty} p_{ik}^{(n-1)} p_{kj} = \sum_{k \in I} \pi_k p_{kj}. \]

**Example 3.1.4** Consider the two-state Markov chain with transition matrix
\[ P = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix}. \]
Ignore the trivial case \( \alpha = \beta = 0 \) and \( \alpha = \beta = 1 \). Then, as we saw before
\[ P^n \to \begin{pmatrix} \frac{\alpha}{\alpha+\beta} & \frac{\beta}{\alpha+\beta} \\ \frac{\alpha}{\alpha+\beta} & \frac{\beta}{\alpha+\beta} \end{pmatrix} \]
as \( n \to \infty \), so, by the previous theorem, the distribution \( \left( \frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta} \right) \) must be invariant. Of course we can solve
\[ \begin{pmatrix} \lambda_1, & \lambda_2 \end{pmatrix} \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix} = \begin{pmatrix} \lambda_1, & \lambda_2 \end{pmatrix}, \]
and we obtain
\[ \begin{pmatrix} \lambda_1, & \lambda_2 \end{pmatrix} = \begin{pmatrix} \frac{\beta}{\alpha+\beta}, & \frac{\alpha}{\alpha+\beta} \end{pmatrix}. \]
Note also that, for the case \( \alpha = \beta = 1 \) we do not have convergence of \( P^n \).

**Example 3.1.5** Consider the three-state chain with matrix
\[ P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}. \]
Then, since the eigenvalues are \( 1, i/2, -i/2 \), we deduce that
\[ p_{11}^{(n)} = a + b \left( \frac{i}{2} \right)^n + c \left( -\frac{i}{2} \right)^n, \]
for some constants \( a, b \) and \( c \). Since \( p_{11}^{(n)} \) is real and
\[ \left( \pm \frac{i}{2} \right)^n = \left( \frac{1}{2} \right)^n e^{\pm i n \pi/2} = \left( \frac{1}{2} \right)^n \left( \cos \frac{n\pi}{2} \pm i \sin \frac{n\pi}{2} \right), \]
then we can write
\[ p_{11}^{(n)} = \alpha + \left(\frac{1}{2}\right)^n \left( \beta \cos \frac{n\pi}{2} + \gamma \sin \frac{n\pi}{2} \right), \]
for constants \( \alpha, \beta \) and \( \gamma \) that can be determined by the boundary conditions
\[
\begin{align*}
1 &= p_{11}^{(0)} = \alpha + \beta \\
0 &= p_{11}^{(1)} = \alpha + \frac{1}{2} \gamma \cos \frac{n\pi}{2} + \frac{1}{4} \beta.
\end{align*}
\]
So, \( \alpha = 1/5, \beta = 4/5, \gamma = -2/5 \) and
\[
p_{11}^{(n)} = \frac{1}{5} + \left(\frac{1}{2}\right)^n \left( \frac{4}{5} \cos \frac{n\pi}{2} - \frac{2}{5} \sin \frac{n\pi}{2} \right). 
\]
Then to find the invariant distribution we can write down the components of the vector equation
\[
\begin{align*}
\pi_1 &= \frac{1}{2} \pi_3 \\
\pi_2 &= \pi_1 + \frac{1}{2} \pi_2 \\
\pi_3 &= \frac{1}{2} \pi_2 + \frac{1}{2} \pi_3.
\end{align*}
\]
One equation is redundant but we have the additional condition
\[ \pi_1 + \pi_2 + \pi_3 = 1 \]
and the solution is \( \pi = (1/5, 2/5, 2/5) \). Note that
\[
p_{11}^{(n)} \to \frac{1}{5} \text{ as } n \to \infty,
\]
so this confirm the previous theorem and also we could have used the result to identify \( \alpha = 1/5 \) instead of working out \( p_{11}^{(2)} \).

For a fixed state \( k \), consider for each \( i \) the expected time spent in \( i \) between visits to \( k \):
\[
\gamma_i^k := E_k \left( \sum_{n=0}^{\tau^{(k)}-1} 1_{\{X_n=i\}} \right).
\]

**Theorem 3.1.7** Let \( P \) be irreducible and recurrent. Then
\[
\begin{align*}
(i) \quad &\gamma_k^k = 1; \\
(ii) \quad &\gamma := (\gamma_i^k, i \in I) \text{ satisfies } \gamma^k P = \gamma^k; \\
(iii) \quad &0 < \gamma_i^k < \infty \text{ for all } i \in I.
\end{align*}
\]
Proof. (i) is obvious. (ii) Since $P$ is recurrent, under $P_k$, $\tau^{(k)} < \infty$ and $X_0 = X_{\tau^{(k)}} = k$

\[
\gamma^k_j = E_k \left( \sum_{n=1}^{\tau^{(k)}} 1_{\{X_n = j\}} \right) = E_k \left( \sum_{n=1}^{\infty} 1_{\{X_n = j\}} 1_{\{n \leq \tau^{(k)}\}} \right)
\]

\[
= \sum_{n=1}^{\infty} P_k(X_n = j, n \leq \tau^{(k)}) = \sum_{i \in I} \sum_{n=1}^{\infty} P_k(X_n = i, X_{n-1} = j, n \leq \tau^{(k)})
\]

\[
= \sum_{i \in I} \sum_{k=1}^{\infty} P_k(X_n = j | X_{n-1} = i, n \leq \tau^{(k)}) P_k(X_{n-1} = i, n \leq \tau^{(k)})
\]

\[
= \sum_{i \in I} \sum_{k=1}^{\infty} p_{ij} \left( \sum_{n=1}^{\tau^{(k)}} 1_{\{X_n = i\}} \right)
\]

\[
= \sum_{i \in I} p_{ij} \gamma^k_i.
\]

(iii) Fixed $i \in I$, since $P$ is irreducible, there exist $m, n \geq 0$ such that $p^{(n)}_{ki} > 0$ and $p^{(m)}_{ik} > 0$, then by (i) and (ii) $\gamma^k_i > \gamma^k_k p^{(n)}_{ki} > 0$ and $\gamma^k_i > \gamma^k_k p^{(m)}_{ik}$. ■

Theorem 3.1.8 Let $P$ be irreducible and let $\lambda$ be an invariant measure for $P$ with $\lambda_k = 1$. Then $\lambda \geq \gamma^k$. If in addition $P$ is recurrent, then $\lambda = \gamma^k$.

Proof.

\[
\lambda_j = \sum_{i_0 \in I} \lambda_{i_0} p_{i_0 j} = \sum_{i_0 \neq k} \lambda_{i_0} p_{i_0 j} + p_{kj}
\]

\[
= \sum_{i_0 \neq k} \sum_{i_1 \neq k} \lambda_{i_0} p_{i_0 i_1} p_{i_1 j} + p_{kj} + \sum_{i_0 \neq k} p_{ki_0} p_{i_0 j}
\]

\[
= \cdots = \sum_{i_0 \neq k} \sum_{i_1 \neq k} \cdots \sum_{i_n \neq k} \lambda_{i_0} p_{i_0 i_1} \cdots p_{i_n j}
\]

\[
+ p_{kj} + \sum_{i_0 \neq k} p_{ki_0} p_{i_0 j} + \cdots + \sum_{i_0 \neq k} \sum_{i_1 \neq k} \cdots \sum_{i_{n-1} \neq k} p_{ki_{n-1}} \cdots p_{i_0 j}
\]

\[
\geq p_{kj} + \sum_{i_0 \neq k} p_{ki_0} p_{i_0 j} + \cdots + \sum_{i_0 \neq k} \sum_{i_1 \neq k} \cdots \sum_{i_{n-1} \neq k} p_{ki_{n-1}} \cdots p_{i_0 j}
\]

\[
\to \gamma^k_j \text{ as } n \text{ goes to infinity.}
\]

Since $P$ is recurrent $\gamma$ is invariant and $\mu := \lambda - \gamma$ will be also invariant, now since it is irreducible, for any fixed $i$, there will be $n \geq 0$ such that $p^{(n)}_{ik} > 0$ and

\[
0 = \mu_k = \sum_j \mu_j p^{(n)}_{jk} = \mu_i p^{(n)}_{ik},
\]

so $\mu_i = 0$. ■
A recurrent state is positive recurrent if the expected return time
\[ m_i := E_i(\tau(i)) \]
is finite, otherwise is called null recurrent.

**Remark 3.1.3** Note that
\[ m_i = E_i \left( \sum_{n=0}^{\tau(i)-1} \sum_{j \neq i} 1_{\{X_n = j\}} \right) + 1 = \sum_{j \in I} \gamma_j^i \]

**Theorem 3.1.9** Let \( P \) be irreducible. Then the following are equivalent:

(i) every state is positive recurrent;
(ii) some state \( i \) is positive recurrent;
(iii) \( P \) has an invariant distribution, \( \pi \) say.

Moreover, when (iii) holds we have \( m_i = 1/\pi_i \), for all \( i \).

**Proof.** (i)\(\Rightarrow\)(ii) is evident. (ii)\(\Rightarrow\)(iii). If \( i \) is recurrent then \( P \) is recurrent and by the Theorem 3.1.7 \( \gamma^i \) is an invariant measure. But
\[ \sum_{j \in I} \gamma_j^i = m_i, \]
and \( m_i < \infty \), so \( \pi_i := \gamma_j^i/m_i \) defines an invariant distribution. (iii)\(\Rightarrow\)(i).

Takes any state \( k \), since \( P \) is irreducible there exist \( n \geq 1 \) such that \( \pi_k^{(n)} \sum_{j \in I} \pi_j \pi_{ik}^{(n)} > 0 \). Set \( \lambda_i = \pi_i/\pi_k \), then, by the previous theorem \( \lambda \geq \gamma^k \) and
\[ m_k = \sum_{i \in I} \gamma_i^k \leq \sum_{i \in I} \frac{\pi_i}{\pi_k} = \frac{1}{\pi_k} < \infty \]
and \( k \) is positive recurrent. Note finally that, since \( P \) is recurrent, by the previous theorem the above inequality is in fact an equality. \( \blacksquare \)

**Theorem 3.1.10** If the state space is finite there is at least one stationary distribution.

**Proof.** We can assume the chain is irreducible, then since it is finite, every state is positive recurrent, in fact since
\[ m_k = \sum_{i \in I} \gamma_i^k \]
and \( I \) is finite it is sufficient to see that \( \gamma_i^k < \infty \), but this follows from Theorem 3.1.7. \( \blacksquare \)
3.1.6 Limiting behaviour

Let us call a state \( i \) is aperiodic if \( p_{ii}^{(n)} > 0 \) for all sufficiently large \( n \). It can be shown that \( i \) is aperiodic iff the set \( \{ n \geq 0, p_{ii}^{(n)} > 0 \} \) has no common divisor other than 1. It can be also shown that it is a class property.

**Theorem 3.1.11** Let \( P \) be irreducible and aperiodic and suposse that \( P \) has an invariant distribution \( \pi \) then
\[
p_{ij}^{(n)} \to \pi_j \quad \text{as } n \to \infty \quad \text{for all } i, j.
\]

3.1.7 Ergodic theorem

Ergodic theory concerns the limiting behaviour of averages over time. We shall prove a theorem which identifies for Markov chains the long-run proportion of time spent in each state.

Denote by \( V_i(n) \) the number of visits to \( i \) before \( n \):
\[
V_i(n) := \sum_{k=0}^{n-1} 1_{\{X_k = i\}}.
\]

**Theorem 3.1.12** Let \( P \) irreducible. Then
\[
\frac{V_i(n)}{n} \to \frac{1}{m_i} \quad \text{a.s. as } n \to \infty.
\]

Moreover, in the positive recurrent case, for any bounded function \( f : I \to \mathbb{R} \) we have
\[
\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) \to \sum_{i \in I} \pi_i f(i), \quad \text{a.s as } n \to \infty.
\]
where \( \pi \) is the unique invariant distribution.

**Proof.** Without loss of generality we assume that we start at state \( i \). If \( P \) is trasient then \( V_i(n) \) is bounded by the total number of visits \( V_i \) so then
\[
\frac{V_i(n)}{n} \leq \frac{V_i}{n} \to 0 = \frac{1}{m_i}
\]
Suposse that \( P \) is recurrent. Let \( \delta_r^{(i)} \) the length of the \( r \) excursion to \( i \). Then by the strong Markov property \( \delta_r^{(i)} \) are i.i.d. with \( E(\delta_r^{(i)}) = m_i \), and
\[
\frac{\delta_1^{(i)} + \ldots + \delta_{V_i(n) - 1}^{(i)}}{V_i(n)} \leq \frac{n}{V_i(n)} \leq \frac{\delta_1^{(i)} + \ldots + \delta_{V_i(n)}^{(i)}}{V_i(n)}
\]
and we can apply the strong law of large numbers.
For the second part, we can assume w.l.o.g. that $|f| \leq 1$ then for any $J \subseteq I$

\[
\left| \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) - \pi_i f(i) \right| = \left| \sum_{i \in I} \left( \frac{V_i(n)}{n} - \pi_i \right) f(i) \right|
\leq \sum_{i \in J} \left| \frac{V_i(n)}{n} - \pi_i \right| + \sum_{i \notin J} \left| \frac{V_i(n)}{n} - \pi_i \right|
\leq \sum_{i \in J} \left| \frac{V_i(n)}{n} - \pi_i \right| + \sum_{i \notin J} \left( \frac{V_i(n)}{n} + \pi_i \right)
\leq 2 \sum_{i \in J} \left| \frac{V_i(n)}{n} - \pi_i \right| + 2 \sum_{i \notin J} \pi_i,
\]

but we saw that $\frac{V_i(n)}{n} \to \pi_i$ as $n \to \infty$ for all $i$.

Then, given $\varepsilon > 0$ we can take $J$ such that

\[
\sum_{i \notin J} \pi_i < \frac{\varepsilon}{4},
\]

and $N(\omega)$ such that for all $n \geq N(\omega)$

\[
\sum_{i \in J} \left| \frac{V_i(n)}{n} - \pi_i \right| < \frac{\varepsilon}{4}.
\]

\section{3.2 Poisson process}

\subsection{3.2.1 Exponential distribution}

\textbf{Definition 3.2.1} We say that $T : \Omega \to [0, \infty]$ follows an exponential distribution with parameter $\lambda$ (we shall write it $E(\lambda)$ for short) if

\[f_T(t) = \lambda e^{-\lambda t} 1_{[0, \infty]}(t).
\]

It is easy to see the following properties:

- $E(T) = 1/\lambda$, $\text{var}(T) = 1/\lambda^2$.
- $T/\mu \sim E(\mu)$.
- $P(T > t + s | T > t) = P(T > s)$ (lack of memory).
- If $S \sim E(\lambda)$ and $T \sim E(\mu)$, then $\min(S, T) \sim E(\lambda + \mu)$ and $P(S < T) = \frac{\lambda}{\lambda + \mu}$.
3.2. POISSON PROCESS

• If \( t_1, t_2, \ldots, t_n \) are independent \( E(\lambda) \) then \( T_n := t_1 + t_2 + \ldots + t_n \sim \text{gamma}(n, \lambda) \).

**Theorem 3.2.1** Let \( T_1, T_2, \ldots \) be a sequence of independent random variables with \( T_n = E(\lambda_n) \) and \( 0 < \lambda_n < \infty \) for all \( n \).

(i) If \( \sum_{n=1}^{\infty} \frac{1}{\lambda_n} < \infty \), then \( P(\sum_{n=1}^{\infty} T_n < \infty) = 1 \).

(ii) If \( \sum_{n=1}^{\infty} \frac{1}{\lambda_n} = \infty \), then \( P(\sum_{n=1}^{\infty} T_n = \infty) = 1 \).

**Proof.** (i) Suppose \( \sum_{n=1}^{\infty} \frac{1}{\lambda_n} < \infty \). Then by the monotone convergence theorem
\[
E\left( \sum_{n=1}^{\infty} T_n \right) = \sum_{n=1}^{\infty} E(T_n) = \sum_{n=1}^{\infty} \frac{1}{\lambda_n} < \infty,
\]
so \( P(\sum_{n=1}^{\infty} T_n < \infty) = 1 \).

(ii) Suppose instead that \( \sum_{n=1}^{\infty} \frac{1}{\lambda_n} = \infty \). Then \( \Pi_{n=1}^{\infty} (1 + \frac{1}{\lambda_n}) = \infty \), by the monotone convergence theorem and the independence
\[
E\left( \exp\left\{ -\sum_{n=1}^{\infty} T_n \right\} \right) = \Pi_{n=1}^{\infty} E(\exp\{ -T_n \}) = \Pi_{n=1}^{\infty} (1 + \frac{1}{\lambda_n})^{-1} = 0,
\]
so \( P(\sum_{n=1}^{\infty} T_n = \infty) = 1 \).

**Theorem 3.2.2** Let \( I \) be a countable set and let \( T_k, k \in I \), be independent random variables with \( T_k \sim E(q_k) \) and \( q := \sum_{k \in I} q_k < \infty \). Set \( T = \inf_k T_k \). Then this minimum is attained at a unique random value \( K \) of \( k \), with probability 1. Moreover \( T \) and \( K \) are independent, with \( T \sim E(q) \) and \( P(K = k) = \frac{q_k}{q} \).

**Proof.**
\[
P(K = k, T \geq t) = P(T_k \geq t, T_j > T_k \text{ for all } j \neq k)
\]
\[
= \int_{t}^{\infty} q_k e^{-q_k s} P(T_j > s \text{ for all } j \neq k) ds
\]
\[
= \int_{t}^{\infty} q_k e^{-q_k s} \Pi_{j \neq k} e^{-q_j s} ds = \frac{q_k}{q} e^{-qt},
\]
Hence \( P(K = k, \text{ for some } k) = 1 \), and \( K \) and \( T \) have the claimed joint distribution.

3.2.2 Poisson process

**Definition 3.2.2** Let \( t_1, t_2, \ldots \) of independent \( E(\lambda) \) random variables, let \( T_n = t_1 + t_2 + \ldots + t_n \), for \( n \geq 1 \), \( T_0 = 0 \) and define \( N_s = \max\{n, T_n \leq s\} \). Then, the process \((N_s)_{s \geq 0}\) is named a Poisson process with parameter (or rate) \( \lambda \).
Remark 3.2.1 We think of the $t_n$ as times between arrivals of customers at a bank. Then $N_s$ is the number of arrivals by the time $s$.

Lemma 3.2.1 $N_s \sim \text{Poisson}(\lambda s)$

Proof. $P(N_s = n) = P(T_n \leq s < T_{n+1})$.

$$T_{n+1} = t_{n+1} + t_n,$$

and $t_{n+1}$ is independent of $T_n$, then

$$P(T_n \leq s < T_{n+1}) = \int_s^0 \int_s^\infty f_{T_n}(t)f_{t_{n+1}}(u-t)dudt$$

$$= \int_0^s \int_s^{\infty} \frac{1}{(n-1)!}\lambda^n t^{n-1}e^{-\lambda t} \lambda e^{-\lambda(u-t)}dudt$$

$$= \frac{\lambda^n}{(n-1)!}e^{-\lambda s} \int_0^s t^{n-1}dt = e^{-\lambda s} (\lambda s)^n.$$  

Lemma 3.2.2 $N_{t+s} - N_s, \ t \geq 0$ is a Poisson process with rate $\lambda$ independent of $(N_r)_{0 \leq r \leq s}$.

Proof. Assume that for the time $s$ there were four arrivals and that $T_4 = u_4$. Then

$$P(t_5 > s - u_4 + t|t_5 > s - u_4) = P(t_5 > t) = e^{-\lambda t}.$$ 

So, the first arrival after $s$ is $E(\lambda)$ and independent of $T_1, T_2, T_3, T_4$. It is clear that $t_5, t_6, \ldots$ are independent of $T_1, T_2, T_3, T_4$. ■

Lemma 3.2.3 $(N_t)_{t \geq 0}$ has independent increments.

Theorem 3.2.3 If $(N_t)_{t \geq 0}$ is a Poisson process

(i) $N_0 = 0$

(ii) $N_{t+s} - N_s \sim \text{Poisson}(\lambda t)$

(iii) $(N_t)_{t \geq 0}$ has independent increments.

Conversely, if (i) (ii) and (iii) hold and the process is right continuous, then $(N_t)_{t \geq 0}$ is a Poisson process.

Proof. (The converse)

$$P(T_1 > t) = P(N_t = 0) = e^{-\lambda t}.$$ 

$$P(t_2 > t|t_1 \leq s) = P(N_{t+s} - N_s = 0|N_r \leq 1, 0 \leq r \leq s) = P(N_{t+s} - N_s = 0) = e^{-\lambda t},$$

so $t_2 \sim E(\lambda)$ independent of $t_1$. ■
3.2. POISSON PROCESS

Theorem 3.2.4 \((N_t)_{t \geq 0}\) is a Poisson process iff

(i) \(N_0 = 0\),

(ii) \(N_t\) is an increasing, right continuous, integer-valued process with independent increments,

(ii) \(P(N_{t+h} - N_t = 0) = 1 - \lambda h + o(h), \ P(N_{t+h} - N_t = 1) = \lambda h + o(h)\), as \(h \downarrow 0\) uniformly in \(t\).

Proof. Suppose that \((N_t)_{t \geq 0}\) is a Poisson process. Then

\[
P(N_{t+h} - N_t \geq 1) = P(N_h \geq 1) = 1 - e^{-\lambda h} = \lambda h + o(h),
\]

\[
P(N_{t+h} - N_t \geq 2) = P(N_h \geq 2) = P(T_1 + T_2 \leq h) \leq P(T_1 \leq h, T_2 \leq h)
\]

\[
= P(T_1 \leq h)P(T_2 \leq h) = (1 - e^{-\lambda h})^2 = o(h).
\]

If (i) (ii) and (iii) hold, for \(i = 2, 3, \ldots\), we have \(P(N_{t+h} - N_t = i) = o(h)\) as \(h \downarrow 0\), uniformly in \(t\). Set \(p_j(t) = P(N_t = j)\). Then, for \(j = 1, 2, \ldots\),

\[
p_j(t+h) = P(N_{t+h} = j) = \sum_{i=0}^{j} P(N_{t+h} - N_t = i)P(N_t = j - i)
\]

\[
= (1 - \lambda h + o(h))p_j(t) + (\lambda h + o(h))p_{j-1}(t) + o(h).
\]

So

\[
\frac{p_j(t+h) - p_j(t)}{h} = -\lambda p_j(t) + \lambda p_{j-1}(t) + O(h),
\]

since this estimate is uniform in \(t\) we can put \(t = s - h\) to obtain

\[
\frac{p_j(s) - p_j(s-h)}{h} = -\lambda p_j(s-h) + \lambda p_{j-1}(s-h) + O(h),
\]

now letting \(h \downarrow 0\) we obtain that \(p_j(t)\) are differentiable and satisfies the differential equation

\[
p'_j(t) = -\lambda p_j(t) + \lambda p_{j-1}(t).
\]

Analogously we can see that

\[
p'_0(t) = -\lambda p_0(t),
\]

and since

\[
p_0(0) = 1, \ p_j(0) = 0 \text{ for } j = 1, 2, \ldots
\]

we have that

\[
p_j(t) = e^{-\lambda t} \frac{(\lambda t)^j}{j!} \text{ for } j = 0, 1, 2, \ldots
\]
3.3 Continuous-time Markov chains

3.3.1 Definition and examples

**Definition 3.3.1** $(X_t)_{t \geq 0}$ is a Markov chain if for any $0 \leq s_0 < s_1 \cdots < s_n < s$ and possible states $i_0, ..., i_n, i, j$ we have

\[
P(X_{t+s} = j | X_s = i, X_{s_n} = i_n, ..., X_{s_0} = i_0) = P(X_t = j | X_0 = i).
\]

**Example 3.3.1** Let $(N_t)_{t \geq 0}$ be a Poisson process with rate $\lambda$ and let $(Y_n)_{n \geq 0}$ be a discrete time Markov chain with transition probabilities $\pi_{ij}$ independent of $(N_t)_{t \geq 0}$. Then $X_t = Y_{N_t}$ is a continuous-time Markov chain. Intuitively, this follows from the memoryless property of the exponential distribution: if $X_s = i$, then, independently of what happened in the past, the time to the next jump will be exponentially distributed with rate $\lambda$ and will go to state $j$ with probability $p_{ij}$. If we write $p_{ij}(t) = P(X_t = j | X_0 = i)$, we have

\[
p_{ij}(t) = \frac{P(X_t = j, X_0 = i)}{P(X_0 = i)} = \frac{P(Y_{N_t} = j, Y_0 = i)}{P(Y_0 = i)} = \sum_{k=0}^{\infty} \frac{P(Y_k = j, Y_0 = i)}{P(Y_0 = i)} P(N_t = k) = \sum_{k=0}^{\infty} \pi_{ij} \frac{e^{-\lambda t} (\lambda t)^k}{k!} = \sum_{k=0}^{\infty} \frac{e^{-\lambda t} (t \pi_{ij})^k}{k!} = e^{t \lambda \Pi - \lambda I} |_{ij} = e^{tQ}_{ij}.
\]

Where $Q = \lambda (\Pi - I)$.

**Chapman-Kolmogorov equation**

**Proposition 3.3.1**

\[
\sum_{k \in I} p_{ik}(s)p_{kj}(t) = p_{ij}(t+s). \tag{3.3}
\]

**Proof.**

\[
p_{ij}(t+s) = P(X_{t+s} = j | X_0 = i) = \sum_{k \in I} P(X_{t+s} = j, X_s = k | X_0 = i) = \sum_{k \in I} P(X_{t+s} = j | X_s = k) P(X_s = k | X_0 = i).
\]

\[\blacksquare\]
(3.3) shows that if we know that transition probability for \( t < t_0 \) for any \( t_0 > 0 \) we know it for all \( t \). This observation suggests that the transition probabilities \( p_{ij}(t) \) can be determined from their derivatives at 0:

\[
q_{ij} = \lim_{h \to 0} \frac{p_{ij}(h)}{h}, \ j \neq i.
\]

If this limit exists we will call \( q_{ij} \) the jump rate from \( i \) to \( j \).

For Example 3.3.1 \( q_{ij} = \lambda \pi_{ij} \).

Example 3.3.1 is atypical. There we started with the Markov chain and then defined its rates: In most cases it is much simpler to describe the system by writing down its transition rates \( q_{ij} \) for \( j \neq i \), which describe the rates at which jumps are made from \( i \) to \( j \).

Example 3.3.2 Poisson process

\[
q_{n,n+1} = \lambda \text{ for all } n \geq 0.
\]

Example 3.3.3 M/M/s queue. Imagine a bank with \( s \) sellers that serve customers who queue in a single line if all of the servers are busy. We imagine that customers arrive at times of a Poisson process with rate \( \lambda \), and that each service time is an independent exponential with rate \( \mu \). As in the previous example \( q_{n,n+1} = \lambda \). To model the departures we let

\[
q_{n,n-1} = \begin{cases} 
 n\mu & 0 \leq n \leq s \\
 s\mu & n \geq \mu
\end{cases}
\]

3.3.2 Computing the transition probability

**Theorem 3.3.1** Let \( Q \) be matrix on a finite set \( I \). Set \( P(t) = e^{tQ} \). Then \( (P(t), t \geq 0) \) has the following properties:

(i) \( P(s + t) = P(s)P(t) \) for all \( s, t \) (semigroup property)

(ii) \( (P(t), t \geq 0) \) is the unique solution to the forward equation

\[
\frac{d}{dt} P(t) = P(t)Q, \quad P(0) = I
\]

(iii) \( (P(t), t \geq 0) \) is the unique solution to the backward equation

\[
\frac{d}{dt} P(t) = QP(t), \quad P(0) = I
\]

(iv) for \( k = 0, 1, 2, \ldots \), we have

\[
\left( \frac{d}{dt} \right)^k |_{t=0} P(t) = Q^k.
\]

**Proof.** For any \( s, t \in \mathbb{R}, sQ \) and \( tQ \) commute, so

\[
e^{sQ}e^{tQ} = e^{(s+t)Q}.
\]

Proving the semigroup property. In fact if \( Q_1 \) and \( Q_2 \) commute

\[
e^{(Q_1 + Q_2)} = \sum_{n=0}^{\infty} \frac{(Q_1 + Q_2)^n}{n!} = \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} Q_1^k Q_2^{n-k}
\]

\[
= \sum_{k=0}^{\infty} \frac{Q_1^k}{k!} \sum_{n=k}^{\infty} \frac{Q_2^{n-k}}{(n-k)!} = e^{Q_1}e^{Q_2}.
\]
The matrix-value power series
\[ P(t) = \sum_{k=0}^{\infty} \frac{(tQ)^k}{k!} \]
has infinite radius of convergence. So each component is differentiable with
derivative given by term-by-term differentiation:
\[ P'(t) = \sum_{k=1}^{\infty} \frac{t^{k-1}Q^k}{(k-1)!} = P(t)Q = QP(t). \]
Hence \( P(t) \) satisfies the forward and backward equations. By repeated term-
by-term differentiation we obtain (iv). To show uniqueness, assume that \( M(t) \)
is another solution of the forward equation, then
\[
\frac{d}{dt} (M(t)e^{-tQ}) = \frac{d}{dt} (M(t))e^{-tQ} + M(t) \frac{d}{dt} (e^{-tQ})
= M(t)Qe^{-tQ} + M(t)(-Q)e^{-tQ} = 0,
\]
so \( M(t)e^{-tQ} \) is constant, and so \( M(t) = e^{tQ}. \) Similarly for the backward equa-
tion. □

**Example 3.3.4** Let
\[ Q = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -1 & 0 \\ 2 & 1 & -3 \end{pmatrix}, \]
\[ \det(Q - \lambda I) = \lambda(\lambda + 2)(\lambda + 4). \]
Then, \( Q \) has eigenvalues 0, -2, -4. Then there is an invertible matrix \( U \) such
that
\[ Q = U \begin{pmatrix} 0 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -4 \end{pmatrix} U^{-1}, \]
so
\[ e^{tQ} = \sum_{k=0}^{\infty} \frac{(tQ)^k}{k!} = U \sum_{k=0}^{\infty} \frac{1}{k!} \begin{pmatrix} 0 & 0 & 0 \\ 0 & (-2t)^k & 0 \\ 0 & 0 & (-4t)^k \end{pmatrix} U^{-1} \]
\[ = U \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{-2t} & 0 \\ 0 & 0 & e^{-4t} \end{pmatrix} U^{-1}. \]
To determine the conditions we use
\[ P(0) = I, \quad P'(0) = Q, \quad P''(0) = Q^2. \]
Theorem 3.3.2 Consider a continuous-time Markov chain, with a finite set of states. Assume that as $h \downarrow 0$ we have that
\[ p_{ij}(h) = \delta_{ij} + q_{ij}h + o(h), \text{ for all } i, j \in I, \]
uniformly in $t$, then \( \{p_{ij}(t), i, j \in I, t \geq 0\} \) is the solution of the forward equation
\[ p'(t) = p(t)Q, \quad p(0) = I, \]
where \( (p(t))_{ij} = p_{ij}(t), \quad Q_{ij} = q_{ij} \).

**Proof.** Similar to the Theorem 3.2.4:
\[ p_{ij}(t + h) = \sum_{k \in I} p_{ik}(t)(\delta_{kj} + q_{kj}h + o(h)). \]
Since $I$ is finite,
\[ \frac{p_{ij}(t + h) - p_{ij}(t)}{h} = \sum_{k \in I} p_{ik}(t)q_{kj} + O(h). \]

**Remark 3.3.1** Note that in the previous theorem $q_{ii} = -\sum_{j \neq i} q_{ij}$. By Theorem 3.3.1 \( p_{ij}(t), i, j \in I, t \geq 0 \) is also the solution of the backward equation
\[ p'(t) = Qp(t), \quad p(0) = I, \]
where \( (p(t))_{ij} = p_{ij}(t), \quad Q_{ij} = q_{ij} \).

**Remark 3.3.2** We have similar results for the case of infinite state-space. But in this case we have to look for a minimal non-negative solution of the backward and forward equations.

**Example 3.3.5** Consider a two-state chain, where, for concreteness $I = \{1, 2\}$. In this case we have to specify only two rates $q_{12} = \lambda$ and $q_{21} = \mu$, so
\[ Q = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}. \]

Writing the backward equation in matrix form
\[ \begin{pmatrix} p'_{11}(t) \\ p'_{21}(t) \end{pmatrix} = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix} \begin{pmatrix} p_{11}(t) \\ p_{21}(t) \end{pmatrix}, \]
to solve this we guess that
\[ p_{11}(t) = \frac{\mu}{\mu + \lambda} + \frac{\lambda}{\mu + \lambda} e^{-(\mu + \lambda)t}, \]
\[ p_{21}(t) = \frac{\mu}{\mu + \lambda} - \frac{\mu}{\mu + \lambda} e^{-(\mu + \lambda)t}. \]
Example 3.3.6 (The Yule process). In this model each particle splits into two at rate $\beta$ so $q_{i,i+1} = \beta_i$. It can be shown that

$$p_{ij}(t) = e^{-\beta t}(1 - e^{-\beta t})^{j-1} \text{ for } j \geq 1.$$  \hspace{1cm} (3.4)

From this, since the chain starting at $i$ is the sum of $i$ copies of the chain starting at $i$ we have

$$p_{ij}(t) = \binom{j+i-1}{j} (e^{-\beta t})^i (1 - e^{-\beta t})^{j-i}.$$  

3.3.3 Jump chain and holding times

We say that a matrix $Q = (q_{ij}, i, j \in I)$ is a rate matrix if

(i) $0 \leq -q_{ii} < \infty$ for all $i$,
(ii) $q_{ij} \geq 0$ for all $i \neq j$,
(iii) $\sum_{j \in I} q_{ij} = 0$ for all $i$.

We shall write $q_i$ or $q(i)$ as an alternative notation for $-q_{ii}$.

From a rate matrix $Q$ we can obtain a stochastic matrix $\Pi$, the jump matrix,

$$\pi_{ij} = \begin{cases} q_{ij}/q_i & \text{if } j \neq i \text{ and } q_i \neq 0 \\ 0 & \text{if } j \neq i \text{ and } q_i = 0 \end{cases}$$

$$\pi_{ii} = \begin{cases} 0 & \text{if } q_i \neq 0 \\ 1 & \text{if } q_i = 0 \end{cases}.$$  

Example 3.3.7

$$Q = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -1 & 0 \\ 2 & 1 & -3 \end{pmatrix},$$

then

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

Definition 3.3.2 A minimal right-continuous process $(X_t)_{t \geq 0}$ on $I$ is a Markov chain with initial distribution $\lambda$ and rate matrix $Q$ if its jump chain $(Y_n)_{n \geq 0}$ is a discrete-time Markov chain $(\lambda, \Pi)$ and if for each $n \geq 1$, conditional on $Y_0, Y_1, ..., Y_{n-1}$, its holding times $S_1, S_2, ..., S_n$ are independent exponential random variables of parameters $q(Y_0), q(Y_1), ..., q(Y_{n-1})$. In such a way that if write the jump times $J_n = S_1 + S_2 + ... + S_n$ $X_{J_n} = Y_n$,

and

$$X_t = \begin{cases} Y_n & \text{if } J_n \leq t < J_{n+1} \text{ for some } n \\ \infty & \text{otherwise (it is minimal)} \end{cases}.$$
3.3. CONTINUOUS-TIME MARKOV CHAINS

Let $X_0 = Y_0$ with distribution $\lambda$, and with and array $(T_n^j, n \geq 1, j \in I)$ of independent exponential random variable of parameter 1. Then, inductively for $n = 0, 1, 2, \ldots$, if $Y_n = i$, we set

\[ S_{n+1}^j = T_n^j / q_{ij}, \quad \text{for } j \neq i, \]
\[ S_{n+1} = \inf_{j \neq i} S_{n+1}^j, \]
\[ Y_{n+1} = \begin{cases} j & \text{if } S_{n+1}^j \leq S_{n+1} < \infty \\
      i & \text{if } S_{n+1} = \infty \end{cases}. \]

Then conditional on $Y_n = i$, $S_{n+1}$ is $E(q_i)$, $q_i = \sum_{j \neq i} q_{ij}$, $Y_{n+1}$ has distribution $(\pi_{ij}, j \in I)$ and $S_{n+1}$ and $Y_{n+1}$ are independent, and independent of $Y_0, Y_1, \ldots, Y_{n-1}$ and $S_1, \ldots, S_n$.

3.3.4 Birth processes

A birth process, say $(X_t)_{t \geq 0}$, is a generalization of a Poisson process in which the parameter $\lambda$ is allowed to depend on the current state of the process. The data for a birth process consist of birth rates $0 \leq q_{i,i+1} < \infty$, where $i = 0, 1, 2, \ldots$. That is $(X_t)_{t \geq 0}$ is a right-continuous process with values in $\mathbb{Z}_+ \cup \{\infty\}$ and conditional on $X_0 = i$, its holding times are $q_{i,i+1}, q_{i+1,i+2}, \ldots$ and its jump chain is given by $Y_{n+1} = i + n$. So

\[ Q = \begin{pmatrix}
-q_{01} & q_{01} & 0 & 0 \\
-q_{12} & q_{12} & 0 & 0 \\
-q_{23} & q_{23} & \ddots & \ddots \\
\end{pmatrix}. \]

Example 3.3.8 (Simple birth process or Yule process) $q_{i,i+1} = \beta_i$. Write $X_t$ the number of individuals at time $t$. Suppose that $X_0 = 1$ and let $J_1$ denote, as above, the time of the first birth, then $J_1 \sim \exp(\beta)$ and

\[ E(X_t) = E(X_t 1_{\{J_1 \leq t\}}) + E(X_t 1_{\{J_1 > t\}}) \]
\[ = \int_0^t \beta e^{-\beta s} E(X_s | J_1 = s) ds + E(1_{\{J_1 > t\}}) \]
\[ = \int_0^t \beta e^{-\beta s} E(X_s | J_1 = s) ds + e^{\beta t}. \]

If we put $\mu(t) = E(X_t)$ then $E(X_t | J_1 = s) = 2\mu(t - s)$: if we have a birth at $s$ thereafter we have like two simple birth process of the same type as $X$. Then

\[ \mu(t) = 2 \int_0^t \beta e^{-\beta s} \mu(t - s) ds + e^{\beta t}, \]
and from here we obtain
\[ \mu'(t) = \beta \mu(t), \]
so the mean population size grows exponentially:
\[ E(X_t) = e^{\beta t}. \]

Note also that we can obtain \( \mu(t) \) from (3.4)
\[
\mu(t) = \sum_{j=1}^{\infty} j p_{1j}(t) = \sum_{j=1}^{\infty} j e^{-\beta t} (1 - e^{-\beta t})^{-1} = \frac{1}{\beta} \frac{d}{dt} \sum_{j=1}^{\infty} (1 - e^{-\beta t})^j = \frac{1}{\beta} \frac{d}{dt} \left( \frac{1 - e^{-\beta t}}{e^{-\beta t}} \right) = e^{\beta t}.
\]

**Definition 3.3.3** If \((J_n)_{n \geq 1}\) \((S_n)_{n \geq 1}\) are the jump times and the holding times of our Markov chain
\[ \zeta = \sup J_n = \sum_{n=1}^{\infty} S_n \]
is called the explosion time.

**Remark 3.3.3** Note that our Markov chains, say \((X_t)_{t \geq 0}\) are minimal processes in the sense that
\[ X_t = \infty \text{ if } t \geq \zeta. \]

**Remark 3.3.4** It is easy to see from Theorem 3.2.1 that the birth process starting from zero explodes if and only if
\[ \sum_{i=0}^{\infty} \frac{1}{q_{i,i+1}} < \infty. \]

**Example 3.3.9** (Non-minimal chain) Consider a birth process \((X_t)_{t \geq 0}\) starting from 0 with rates \(q_{i,i+1} = 2^i\) for \(i \geq 0\). Then by the previous remark the process explodes, we have insisted that \(X_t = \infty\) if \(t \geq \zeta\), where \(\zeta\) is the explosion time. But another obvious possibility is to start the process off again from zero at time \(\zeta\) and do the same for all subsequent explosions. Using the memoryless property of the exponential distribution it can be shown that the new process is a continuous Markov chain in the sense of Definition 3.3.1.

### 3.3.5 Class structure

Hereafter we deal only with minimal chains, those that die after explosion. Then the class structure is simply the discrete-time class structure of the jump chain \((Y_n)_{n \geq 0}\). We say that \(i\) leads to \(j\) and write \(i \rightarrow j\) if
\[ P_t(X_t = i \text{ for some } t \geq 0) > 0. \]
We say \( i \) communicates with \( j \) and write \( i \leftrightarrow j \) if both \( i \rightarrow j \) and \( j \rightarrow i \). The notions of communicating class, closed class, absorbing state and irreducibility are inherited from the jump chain.

**Theorem 3.3.3** For distinct states \( i \) and \( j \) the following are equivalent:

(i) \( i \rightarrow j \);
(ii) \( i \rightarrow j \) for the jump chain;
(iii) \( q_{i_0i_1}q_{i_1i_2} \cdots q_{i_{n-1}i_n} > 0 \) for some states \( i_0, i_1, \ldots, i_n \) with \( i_0 = i \), and \( i_n = j \);
(iv) \( p_{ij}(t) > 0 \) for all \( t > 0 \);
(v) \( p_{ij}(t) > 0 \) for some \( t > 0 \).

### 3.3.6 Hitting times

Let \((X_t)_{t \geq 0}\) be a Markov chain with rate matrix \(Q\). The hitting time of a subset \(A\) of \(I\) is the random variable \(D_A\) defined by

\[
D_A(\omega) = \inf\{t \geq 0, X_t(\omega) \in A\}
\]

with the usual convention that \(\inf \emptyset = \infty\). Since \((X_t)_{t \geq 0}\) is minimal, if \(H_A\) is the hitting time of \(A\) for the jump chain, then

\[
\{H_A < \infty\} = \{D_A < \infty\},
\]

and on this set

\[
D_A = J_{H_A}.
\]

The probability, starting from \(i\), that \((X_t)_{t \geq 0}\) ever hits \(A\) is then

\[
h_i^A = P_i(D_A < \infty) = P_i(H_A < \infty).
\]

**Theorem 3.3.4** The vector of hitting probabilities \(h^A = (h_i^A, i \in I)\) is the minimal non-negative solution to the system of linear equations

\[
\begin{align*}
    h_i^A &= 1 & & \text{for } i \in A \\
    \sum_{j \in I} q_{ij} h_j^A &= 0 & & \text{for } i \notin A
\end{align*}
\]
The average time taken, starting from \( i \), for \( (X_t)_{t \geq 0} \) to reach \( A \) is given by

\[
k_i^A = E_i(D^A)
\]

**Theorem 3.3.5** Assume that \( q_i > 0 \) for all \( i \notin A \). The vector of expected hitting times \( k^A = (k_i^A, i \in A) \) is the minimal non-negative solution to the system of linear equations

\[
\begin{align*}
  k_i^A &= 0 & \text{for } i \in A \\
  -\sum_{j \in I} q_{ij} k_j^A &= 1 & \text{for } i \notin A
\end{align*}
\]

(3.5)

**Proof.** First note that \( k^A \) satisfies (3.5): if \( X_0 = i \in A \) then \( D_A = 0 \), so \( k_i^A = 0 \); if \( X_0 = i \notin A \), by the Markov property

\[
E_i(D^A - J_1 | Y_1 = j) = E_j(D^A),
\]

so

\[
k_i^A = E_i(D^A) = E_i(J_1) + E_i(D^A - J_1)
\]

\[
= E_i(J_1) + \sum_{j \notin i} E_i(D^A - J_1 | Y_1 = j) P_i(Y_1 = j)
\]

\[
= \frac{1}{q_i} + \sum_{j \notin i} E_j(D^A) \pi_{ij} = \frac{1}{q_i} + \sum_{j \notin i} \pi_{ij} k_j^A
\]

\[
= \frac{1}{q_i} + \sum_{j \notin i} \frac{q_j}{q_i} k_j^A,
\]

and then

\[
-q_i k_i^A = \sum_{j \notin i} q_{ij} k_j^A = 1.
\]

Suppose now that \( y = (y_i, i \in I) \) is another non-negative solution to (3.5). Then \( k_i^A = y_i = 0 \) for \( i \in A \). If \( i \notin A \), then

\[
y_i = \frac{1}{q_i} + \sum_{j \notin A} \pi_{ij} y_j = \frac{1}{q_i} + \sum_{j \notin A} \pi_{ij} \left( \frac{1}{q_j} + \sum_{k \notin A} \pi_{jk} y_k \right)
\]

\[
= E_i(S_1) + E_i(S_2 \{H^A \geq 2\}) + \sum_{j \notin A} \sum_{k \notin A} \pi_{ij} \pi_{jk} y_k.
\]

By repeated substitution for \( y \) we obtain

\[
y_i = E_i(S_1) + \cdots + E_i(S_n \{H^A \geq n\}) + \sum_{j_1 \notin A} \cdots \sum_{j_n \notin A} \pi_{ij_1} \cdots \pi_{j_{n-1}j_n} y_{j_n}.
\]

So,

\[
y_i \geq \sum_{m=1}^{n} E_i(S_m \{H^A \geq m\}) = E_i \left( \sum_{m=1}^{n} S_m \right) \xrightarrow{n \to \infty} E_i(D^A) = k_i^A.
\]
3.3. CONTINUOUS-TIME MARKOV CHAINS

3.3.7 Recurrence and trasience

Let \((X_t)_{t \geq 0}\) be Markov chain with rate matrix \(Q\). Recall that \((X_t)_{t \geq 0}\) is minimal. We say a state \(i\) is recurrent if

\[
P_i(\{t \geq 0, X_t = i\} \text{ is unbounded}) = 1.
\]

We say is transient if

\[
P_i(\{t \geq 0, X_t = i\} \text{ is unbounded}) = 0.
\]

**Remark 3.3.5** Note that if \((X_t)_{t \geq 0}\) can explode starting from \(i\) then \(i\) is certainly not recurrent.

**Theorem 3.3.6** We have:

(i) if \(i\) is recurrent for the jump chain \((Y_n)_{n \geq 0}\), then \(i\) is recurrent for \((X_t)_{t \geq 0}\);

(ii) if \(i\) is transient for the jump chain \((Y_n)_{n \geq 0}\), then \(i\) is transient for \((X_t)_{t \geq 0}\);

(iii) every state is either recurrent or transient;

(iv) recurrence and transient are class properties.

Denote by \(T_i\) the first passage time of \((X_t)_{t \geq 0}\) to state \(i\), defined by

\[
T_i(\omega) = \inf\{t \geq J_1(\omega), X_t(\omega) = i\}
\]

**Proof.** (i) Suppose \(i\) is recurrent for the jump chain \((Y_n)_{n \geq 0}\). If \(X_0 = i\) then \((X_t)_{t \geq 0}\) does not explode and \(J_n \to \infty\). Also \(X(J_n) = Y_n = i\) infinitely often, so \(\{t \geq 0, X_t = i\}\) is unbounded, with probability 1.

(ii) Suppose \(i\) is transient for \((Y_n)_{n \geq 0}\). If \(X_0 = i\) then

\[
N = \sup\{n \geq 0, Y_n = i\} < \infty,
\]

so \(\{t \geq 0, X_t = i\}\) is bounded by \(N_{N+1}\) which is finite with probability 1, because \(\{Y_n, n \leq N\}\) cannot include an absorbing state. 

We denote by \(T_i\) the first passage time of \((X_t)_{t \geq 0}\) to state \(i\), defined by

\[
T_i(\omega) = \inf\{t \geq J_1, X_t(\omega) = i\}.
\]

**Theorem 3.3.7** The following dichotomy holds:

(i) if \(q_i = 0\) or \(P_i(T_i < \infty) = 1\), then \(i\) is recurrent and \(\int_0^\infty p_{ii}(t)dt = \infty\);

(ii) if \(q_i > 0\) and \(P_i(T_i < \infty) < 1\), then \(i\) is transient and \(\int_0^\infty p_{ii}(t)dt < \infty\).

**Proof.** If \(q_i = 0\) then \((X_t)_{t \geq 0}\) cannot leave \(i\) and \(p_{ii}(t) = 1\). Suppose \(q_i > 0\). Let \(N_i\) denote the first passage time of the chain \((Y_n)_{n \geq 0}\) for the state \(i\). Then

\[
P(N_i < \infty) = P(T_i < \infty),
\]
so $i$ is recurrent if and only if $P(T_i < \infty) = 1$ by the previous theorem and the result for the jump chain $(Y_n)_{n \geq 0}$.

Write $\pi_{ij}^{(n)}$ for $(\Pi^n)_{ij}$. We shall prove that

$$\int_0^\infty p_{ii}(t)dt = \frac{1}{q_i} \sum_{n=0}^{\infty} \pi_{ij}^{(n)}.$$

$$\int_0^\infty p_{ii}(t)dt = \int_0^\infty (E_i 1_{X_t = i})dt = E_i \left( \int_0^\infty 1_{X_t = i} dt \right)$$

$$= E_i \left( \sum_{n=0}^{\infty} S_{n+1} 1_{Y_n = i} \right)$$

$$= \sum_{n=0}^{\infty} E_i(S_{n+1}|Y_n = i) P_i(Y_n = i) = \frac{1}{q_i} \sum_{n=0}^{\infty} \pi_{ij}^{(n)}.$$

3.3.8 Invariant distributions

We say that $\lambda$ is invariant if

$$\lambda Q = 0.$$

Theorem 3.3.8 Let $Q$ a rate matrix with jump matrix $\Pi$ and let $\lambda$ be a measure. The following are equivalent:

(i) $\lambda$ is invariant;

(ii) $\mu \Pi = \mu$ where $\mu_i = \lambda_i q_i$.

Proof. We have $q_i(\pi_{ij} - \delta_{ij}) = q_{ij}$ for all $i, j$ so

$$\mu (\Pi - I)_{ij} = \sum_{i \in I} \mu_i (\pi_{ij} - \delta_{ij}) = \sum_{i \in I} \lambda_i q_{ij}.$$

Theorem 3.3.9 Suppose that $Q$ is irreducible and recurrent. Then $Q$ has an invariant measure $\lambda$ unique up to scalar multiples.

Recall that a state $i$ is recurrent if $q_i = 0$ or $P_i(T_i < \infty) = 1$. If $q_i = 0$ or the expected return time $m_i = E_i(T_i)$ is finite we say $i$ is positive recurrent otherwise is called null recurrent.

Theorem 3.3.10 Let $Q$ be an irreducible rate matrix. Then the following are equivalent:

(i) every state is positive recurrent;

(ii) some state $i$ is positive recurrent;

(iii) $P$ has an invariant distribution, $\lambda$ say.

Moreover, when (iii) holds we have $m_i = 1/(\lambda_i q_i)$, for all $i$. 
Theorem 3.3.11 Let $Q$ be an irreducible rate matrix, and let $\lambda$ be a measure. Let $s > 0$ be given. The following are equivalent

(i) $\lambda Q = 0$;
(ii) $\lambda P(s) = \lambda$.

3.3.9 Convergence to equilibrium

Theorem 3.3.12 Let $Q$ be an irreducible non-explosive rate matrix with semigroup $P(t)$ having an invariant distribution $\lambda$. Then for all states $i, j$ we have

$p_{ij}(t) \to \lambda_j$ as $t \to \infty$.

Theorem 3.3.13 Let $Q$ be an irreducible rate matrix and let $\nu$ be any distribution. Suppose that $(X_t)_{t \geq 0}$ is Markov $(\nu, Q)$. Then

$P(X_t = j) \to 1/(q_j m_j)$ as $t \to \infty$ for all $j \in I$

where $m_j$ is the expected return time to state $j$. 