Notes on Stochastic Network Modeling (SNM)

Llorenç Cerdà-Alabern
llorenc@ac.upc.edu
Barcelona, October 18, 2020

Contents

Preface

These notes were prepared for the subject *Stochastic Network Modeling (SNM)* from the *Master in Innovation and Research in Informatics (MIRI)*, Universitat Politècnica de Catalunya.

Further reading: The book of Mor [3], Nelson [6] or Trivedi [7] are excellent books on Markov Chains and cover most of the course. The book of Kemeny [4] covers absorbing DTMC.

Part I

Introduction

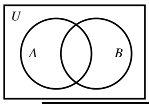
Chapter 1
Probability Review

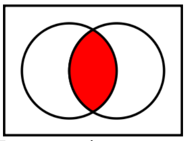
1.1 Ingredients of Probability

- **Random experiment**, e.g. toss a die.
- **Outcome,** ω, e.g. tossing a die can be $\omega = 2$, choosing a fruit can be $\omega =$ orange.
- **Sample space or Universal set,** U, set of all possible outcomes. E.g. tossing a die $U = \{1,2,3,4,5,6\}$.
- **Event,** A, any subset of U (e.g. tossing a die $A = \{1,2,3\}$). We say the event A occurs if the outcome of the experiment $\omega \in A$. U is the **sure event**, and we represent by the empty set $\emptyset$ an **impossible outcome**.

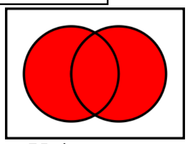
Venn Diagrams

Graphical representation of events

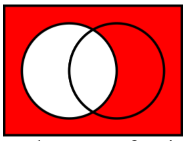




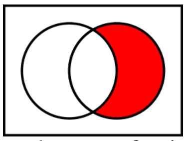
Intersection $A \cap B$



Union $A \cup B$



Complement of A in U
$A^c = U \setminus A$



Complement of A in B
(B minus A)
$A^c \cap B = B \setminus A$

source: wikipedia

Random Variable

- For simplicity it is defined a **random variable (RV),** X as a function that assigns a real number to each outcome in the sample space U, i.e.:

$$X : U \to \mathbb{R} \tag{1.1}$$

- We will represent the experiment by a RV, X, and the possible outcomes by its values. $X = x_i$ **is the outcome** $X(\omega_i) = x_i$.
- Using RVs the sample space is mapped in a subset of $\mathbb{R}$. So, in terms of X, U **is a set of points of** $\mathbb{R}$. **The same for any event.**
- Normally the definition of X **comes naturally from the experiment**, e.g. tossing a die: $X = \{\text{number in the toss}\}$.
- RVs can be **discrete** (e.g. tossing a die) or **continuous** (e.g. waiting time of a packet in a queue).

Probability Measure

- If the sample space U of the RV X is **finite** (**discrete RV**), $U = \{x_1, \cdots x_n\}$, a **probability measure** is an assignment of numbers $P(x_i)$, referred to as **probabilities**, to each **outcome** x_i such that:

$$\begin{aligned} &0 \le P(x_i) \le 1 \\ &P(A) = \sum_{x_i \in A} P(x_i) \\ &P(U) = 1 \end{aligned} \tag{1.2}$$

E.g. tossing a fair die,

$$\begin{aligned} &P(x_i) = 1/6 \\ &P(X \in \{2,4,6\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2} \end{aligned} \tag{1.3}$$

- If the sample space of the RV X is continuous (**continuous RV**), the events are intervals of $\mathbb{R}$. The probability measure is defined by means of the **cumulative distribution function, CDF**:

$$F(x) = P(X \in (-\infty, x]) = P(X \le x) \tag{1.4}$$

- X is called absolutely continuous[1] if there exists the **probability density function, PDF**, such that for any interval $I = \{x \mid a \le x \le b\}$:

$$\int_a^b f(x)\,dx = P(X \in I) = F(b) - F(a) \tag{1.5}$$

Conditional Probability and Bayes Formula

- Given the the sample space U and the **events** $A, B \in U$ with $P(B) > 0$ the **probability of** A **conditioned by** B is defined as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{1.6}$$

NOTE: It's common to use commas to denote set intersection, and write $P(A \cap B)$ **as** $P(A,B)$.

- **Bayes Formula**

$$P(A|B)\,P(B) = P(B|A)\,P(A) \Rightarrow P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} \tag{1.7}$$

Law of total probability

- Let B_i a **partition** of the sample space U ($\cup_i B_i = U$, $B_i \cap B_j = \emptyset, \forall i \ne j$), then

$$P(A) = \sum_i P(A|B_i)\,P(B_i) \tag{1.8}$$

- For **conditional probabilities**:

$$P(A|C) = \sum_i P(A|C \cap B_i)\,P(B_i|C) \tag{1.9}$$

- If C is **independent** of any of the B_i

$$P(A|C) = \sum_i P(A|C \cap B_i)\,P(B_i) \tag{1.10}$$

1.2 Expected value

- Given the discrete $N \in \mathbb{Z}$, respectively continuous $X \in \mathbb{R}$ RV, the **expected value** is:

$$\begin{aligned} \mathrm{E}[N] &= \sum_{k=-\infty}^{\infty} k\,P(N = k) \\ \mathrm{E}[X] &= \int_{-\infty}^{\infty} x\,f(x)\,dx \end{aligned} \tag{1.11}$$

[1] Some special distributions, called singular, do not have a PDF. One example is the Cantor distribution (see Wikipedia).

Example A number $X_1 \in \{1,2,\cdots 6\}$ is obtained tossing a dice. Then, a number $X_2 \in [0,\infty]$ is obtained exponentially distributed with parameter X_1. Compute $f(x_1,x_2)$, $f(x_2)$ and $\mathrm{E}[X_2]$.
Note: Exponential distribution with parameter α:

$$f(x) = \alpha\,\mathrm{e}^{-\alpha x}, x \in [0,\infty], \mathrm{E}[X] = \frac{1}{\alpha}. \tag{1.12}$$

Solution:

$$f(x_1,x_2) = f(x_2|x_1)\,P(x_1) = x_1\,\mathrm{e}^{-x_1 x_2}\,\frac{1}{6}, \begin{cases} x_1 \in \{1,2,\cdots 6\} \\ x_2 \in [0,\infty] \end{cases}$$

$$f(x_2) = \sum_{x_1} f(x_2|x_1)\,P(x_1) = \frac{1}{6}\sum_{n=1}^{6} n\,\mathrm{e}^{-n x_2}, x_2 \in [0,\infty]$$
$$\mathrm{E}[X_2] = \frac{1}{6}\sum_{n=1}^{6}\int_{x_2=0}^{\infty} x_2\,n\,\mathrm{e}^{-n x_2} = \frac{1}{6}\sum_{n=1}^{6}\frac{1}{n} = \frac{49}{120}$$

1.3 Variance

- The amount of dispersion of a RV X with expected value $\mu = \mathrm{E}[X]$ is measured by the **Variance**:

$$\mathrm{Var}(X) = \sigma^2 = \mathrm{E}[(X-\mu)^2] = \mathrm{E}[X^2] - \mu^2 \tag{1.13}$$

- Often it is used the **standard deviation** $\sigma = \sqrt{\mathrm{Var}(X)}$.

1.4 Indicator Function

$$I(A) = \begin{cases} 1, & \text{if } A \text{ occurs,} \\ 0, & \text{otherwise.} \end{cases} \tag{1.14}$$

Therefore:

$$\mathrm{E}[\mathbf{I(A)}] = 0 \times P(I(A)=0) + 1 \times P(I(A)=1) = \mathbf{P(A)} \tag{1.15}$$

1.5 Expected value of non negative RVs

- For **non negative** RVs, $N \geq 0$ discrete and $X \geq 0$ continous:

$$\begin{aligned} \mathrm{E}[N] &= \sum_{k=0}^{\infty} k\,P(N=k) = \sum_{k=0}^{\infty} P(N>k) \\ \mathrm{E}[X] &= \int_0^{\infty} x\,f(x)\,dx = \int_0^{\infty} P(X>x)\,dx = \int_0^{\infty}(1-F(x))\,dx \end{aligned} \tag{1.16}$$

$$\begin{aligned} N &= \sum_{k=0}^{N-1} 1 = \sum_{k=0}^{\infty} I(N>k) \\ X &= \int_0^X dx = \int_0^{\infty} I(X>x)dx \end{aligned} \tag{1.17}$$

and take expectations.

1.6 Wald's Equation

- **Definition**: An positive integer RV $N > 0$ is a **stopping time** of a sequence $X_1, X_2, \cdots$ if the event $N = n$ is independent of $X_{n+1}, X_{n+2}, \cdots$.
  E.g. toss a die until you get 6. Let N be the number of tosses. N does not depend on the values obtained after getting 6.
- **Wald's Equation** If $X_1, X_2, \cdots$ are independent and identically distributed and N is a stopping time:

$$\boxed{\mathrm{E}\left[\sum_{n=1}^{N} X_n\right] = \mathrm{E}[X]\,\mathrm{E}[N]} \tag{1.18}$$

Proof.
$$\begin{aligned} \mathrm{E}\left[\sum_{n=1}^{N} X_n\right] &= \mathrm{E}\left[\sum_{n=1}^{\infty} X_n\,I(n\leq N)\right] = \\ &\sum_{n=1}^{\infty}\mathrm{E}[X_n]\,\mathrm{E}[I(n\leq N)] = \\ &\mathrm{E}[X]\sum_{n=1}^{\infty} P(n\leq N) = \\ &\mathrm{E}[X]\sum_{n=0}^{\infty} P(N>n) = \mathrm{E}[X]\,E[N] \quad \square \end{aligned}$$

1.7 Probability in $\mathbb{R}^k$

If we have a set of k RV $\boldsymbol{X} = (X_1, \cdots X_k)$ taking values in $\mathbb{R}^k$ ($\boldsymbol{X} \in \mathbb{R}^k$), we define the **joint distribution**:

- Discrete RV

$$P(\boldsymbol{x}) = P(x_1, \cdots x_k) = P(X_1 = x_1, \cdots X_k = x_k) \tag{1.19}$$

- Continuos RV:
  - **cumulative distribution function, CDF**:

$$F(\boldsymbol{x}) = F(x_1, \cdots x_k) = P(X_1 \in (-\infty, x_1], \cdots X_k \in (-\infty, x_k]) \tag{1.20}$$

  - with **joint density** function $f(\boldsymbol{x}) = f(x_1, \cdots x_k)$ (if exists):

$$\begin{aligned} F(\boldsymbol{x}) &= F(x_1, \cdots x_k) = \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_k} f(x_1, \cdots x_k)\,dx_k \cdots dx_1 \\ f(\boldsymbol{x}) &= f(x_1, \cdots x_k) = \frac{\partial^k F(x_1, \cdots x_k)}{\partial x_1 \cdots \partial x_k} \end{aligned} \tag{1.21}$$

Marginal distributions Let $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$, where $\boldsymbol{X} \in \mathbb{R}^k, \boldsymbol{X}_1 \in \mathbb{R}^r, \boldsymbol{X}_2 \in \mathbb{R}^{k-r}$, $1 \leq r < k$:

- Discrete RV

$$P(\boldsymbol{x}_2) = \sum_{x_1}\cdots\sum_{x_r} P(\boldsymbol{x}_1, \boldsymbol{x}_2) \tag{1.22}$$

- Continuos RV

$$f(\boldsymbol{x}_2) = \int_{x_1}\cdots\int_{x_r} f(\boldsymbol{x}_1, \boldsymbol{x}_2)\,dx_1 \cdots dx_r \tag{1.23}$$

Independent RV

- Discrete RV

$$\begin{aligned} P(\boldsymbol{x}) &= P(x_1, \cdots x_k) = \\ &P(X_1 = x_1, \cdots X_k = x_k) = \\ &P(X_1 = x_1) \cdots P(X_k = x_k) \end{aligned} \tag{1.24}$$

- Continuos RV

$$\begin{aligned} F(\boldsymbol{x}) &= F(x_1, \cdots x_k) = F_{X_1}(x_1) \cdots F_{X_k}(x_k) \\ f(\boldsymbol{x}) &= f(x_1, \cdots x_k) = f_{X_1}(x_1) \cdots f_{X_k}(x_k) \end{aligned} \tag{1.25}$$

Conditional Distribution

- Let $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$, where $\boldsymbol{X} \in \mathbb{R}^k, \boldsymbol{X}_1 \in \mathbb{R}^r, \boldsymbol{X}_2 \in \mathbb{R}^{k-r}$, the r-dimensional distribuion of $\boldsymbol{X}_1$ **conditioned by** $\boldsymbol{X}_2 = \boldsymbol{x}_2$, $P(\{\boldsymbol{X}_2 = \boldsymbol{x}_2\}) > 0$ is:

$$F(\boldsymbol{X}_1|\boldsymbol{X}_2) = P(\boldsymbol{X}_1 \le \boldsymbol{x}_1|\boldsymbol{X}_2 = \boldsymbol{x}_2) = \frac{P(\boldsymbol{X}_1 \le \boldsymbol{x}_1, \boldsymbol{X}_2 = \boldsymbol{x}_2)}{P(\boldsymbol{X}_2 = \boldsymbol{x}_2)}.$$

If $\boldsymbol{X}$ is **discrete** with probability $P(\boldsymbol{x}_1, \boldsymbol{x}_2)$ or absolutely **continuous** with density $f(\boldsymbol{x}_1, \boldsymbol{x}_2)$:

$$\begin{aligned} P(\boldsymbol{x}_1|\boldsymbol{x}_2) &= \frac{P(\boldsymbol{x}_1, \boldsymbol{x}_2)}{P(\boldsymbol{x}_2)} \\ f(\boldsymbol{x}_1|\boldsymbol{x}_2) &= \frac{f(\boldsymbol{x}_1, \boldsymbol{x}_2)}{f(\boldsymbol{x}_2)} \end{aligned} \tag{1.26}$$

Composition of marginals and conditionals Using the previous formulas we can compute ($\boldsymbol{X}$ can be a mixture of discrete and continuous RV):

Law of total probability

- If $\boldsymbol{x}_1, \boldsymbol{x}_2$ are **discrete** RV: $P(\boldsymbol{x}_2) = \sum_{\boldsymbol{x}_1} P(\boldsymbol{x}_2|\boldsymbol{x}_1)\, P(\boldsymbol{x}_1)$
- If $\boldsymbol{x}_1$ is **discrete** and $\boldsymbol{x}_2$ is **cont.**: $f(\boldsymbol{x}_2) = \sum_{\boldsymbol{x}_1} f(\boldsymbol{x}_2|\boldsymbol{x}_1)\, P(\boldsymbol{x}_1)$
- If $\boldsymbol{x}_1, \boldsymbol{x}_2$ are **cont.**: $f(\boldsymbol{x}_2) = \int_{\boldsymbol{x}_1} f(\boldsymbol{x}_2|\boldsymbol{x}_1)\, f(\boldsymbol{x}_1)\, d\boldsymbol{x}_1$
- If $\boldsymbol{x}_1$ is **cont.** and $\boldsymbol{x}_2$ is **discrete**: $P(\boldsymbol{x}_2) = \int_{\boldsymbol{x}_1} P(\boldsymbol{x}_2|\boldsymbol{x}_1)\, f(\boldsymbol{x}_1)\, d\boldsymbol{x}_1$

Conditional expected value

- Given $X \in \mathbb{R}$, $\boldsymbol{Y} \in \mathbb{R}^k$ with density $f(x, \boldsymbol{y})$:

$$\begin{aligned} \mathrm{E}[X \mid \boldsymbol{Y} = \boldsymbol{y}] &= \int_{\mathbb{R}} x\, f(x|\boldsymbol{y})\, dx \\ \mathrm{E}[X] &= \int_{\mathbb{R}^k} \mathrm{E}[X \mid \boldsymbol{Y} = \boldsymbol{y}]\, f(\boldsymbol{y})\, d\boldsymbol{y} \end{aligned} \tag{1.27}$$

where the **marginal** $f(\boldsymbol{y}) = \int_{x=-\infty}^{\infty} f(x, \boldsymbol{y})\, dx$ and the **conditional** $f(x|\boldsymbol{y}) = f(x, \boldsymbol{y}) / f(\boldsymbol{y})$.
Thus, the law of total probability also applies to expected value, and it is known as **law of total expectation**.

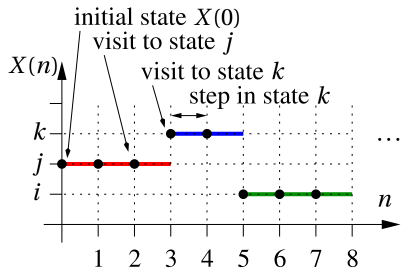
Chapter 2
Stochastic Process (SP)

2.1 Introduction

- **Sequence of RVs** $\{X(t)\}_{t \ge 0}$.
- $X(t)$ is the **state** at time t.
- The **state** $X(t)$ can be **continuous** or **discrete**.
- The **index** can be **continuous** or **discrete**. We shall use n for the **index**, and refer to it as **steps** when it is **discrete**, and t and refer to it as **time** when it is **continuous**.
- We call a possible sequence of states of the SP the **sample function** (or sample path) of the SP.

Sample Path

- Possible evolution (**sample path**) of a **discrete state**, **discrete time** SP $\{X(n)\}_{n \ge 0}$:



- To characterize the stochastic process we would need the distribution and **joint probabilities** of the $\{X(n)\}_{n \ge 0}$ RVs:

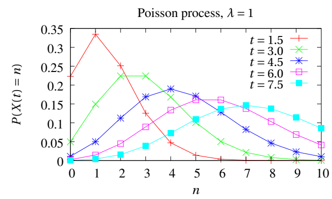
$$P\left(X(n) = i, X(n-1) = k, \cdots X(0) = j\right) \tag{2.1}$$

Example 2: Poisson Process

- It is a discrete state continuous time SP.
- It counts the number of events ocurred in a time interval.
- Often used to build models of other stochastic processes.
- Definition: The number of "events" in any interval of length t, $X(t)$, is **Poisson distributed** with mean λt, i.e.

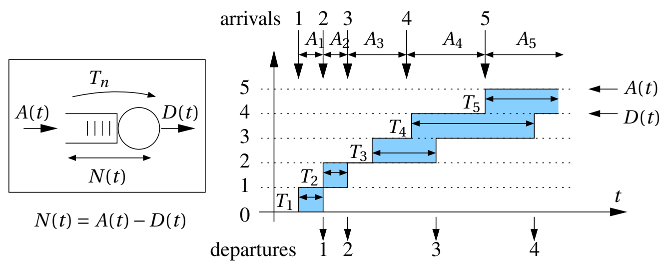
$$\boxed{\begin{aligned} P(X(t+s) - X(s) = n) = P(X(t) - X(0) = n) = \\ P(X(t) = n) = \mathrm{e}^{-\lambda t} \frac{(\lambda t)^n}{n!} \end{aligned}} \tag{2.2}$$

where we assume $X(0) = 0$.



Example 3: Queue with Poisson Arrivals

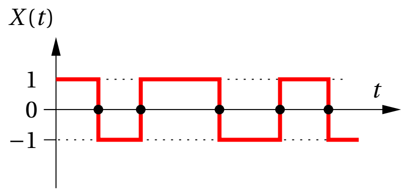
- The queue arrivals, $A(t)$, are modeled as a **Poisson process** with mean λt. Each event model an arrival.



Example 4: Telegraph signal

- The signal is modeled as a **Poisson process** with mean λt such that $X(0) = 1$ or $X(0) = -1$ with equal probability of 1/2 and:

$$X(t) = \begin{cases} 1 & \text{if the number of events in } (0, t] \text{ is even} \\ -1 & \text{if the number of events in } (0, t] \text{ is odd} \end{cases} \tag{2.3}$$



2.2 Analysis of Stochastic Processes

- **Signal Theory**: Normally interested in the **spectral analysis** of the signal. The basic tool is the **Fourier transform** of the **autocorrelation function** of the process (**energy spectral density**). We will not do this analysis.

$$R(t) = \mathrm{E}[X(\tau)\,X(\tau - t)] \qquad \text{autocorrelation}$$

$$F(f) = \mathscr{F}[R(t)] = \int_{-\infty}^{\infty} R(t)\,\mathrm{e}^{-j2\pi f t}\,\mathrm{d}t \qquad \text{Fourier transform (**energy spectral density**)} \tag{2.4}$$

- **Computer Networks**: Normally interested in probabilistic models using **Markov Chains** and **Queueing Theory**.

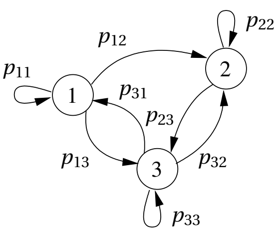
Part II

Discrete Time Markov Chains

Chapter 3 Definition of a DTMC

3.1 State Transition Diagram

- We are interested in a **process that evolve in stages**.
- For the model to be tractable, it is convenient to represent the SP by giving all **possible states** (there may be ∞), and the **possible transitions** between them:



For the model to be consistent: $\sum_{\forall j} p_{ij} = 1$

- Mathematically:

$$p_{ij} = P\big(X(n) = j \mid X(n-1) = i\big) \tag{3.1}$$

3.2 Properties of a DTMC

- The event $X(n) = i$ (at step n the system is in state i) must satisfy (**memoryless property**):

$$P(X(n) = j \mid X(n-1) = i, X(n-2) = k, \cdots) = P(X(n) = j \mid X(n-1) = i) \tag{3.2}$$

- If $P\big(X(n) = j \mid X(n-1) = i\big) = P\big(X(1) = j \mid X(0) = i\big)$ for any n we have an **homogeneous** DTMC. We shall only consider homogeneous DTMC.
- We call **one-step transition probabilities** to:

$$p_{ij} = P\big(X(n) = j \mid X(n-1) = i\big) \tag{3.3}$$

- The SP is called a Markov Process (MP) or Markov Chain (MC) depending on the state being continuous or discrete.

3.3 Transition Matrix

- Transition probabilities:

$$p_{ij} = P\big(X(n) = j \mid X(n-1) = i\big) \tag{3.4}$$

- In matrix form:

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots \\ p_{21} & p_{22} & \cdots \\ \cdots & \cdots & \cdots \end{bmatrix} \tag{3.5}$$

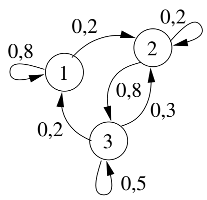
- For the model to be consistent, the probability to move from i to any state must be 1. Mathematically:

$$\sum_{\forall j} p_{ij} = \sum_{\forall j} P\big(X(n) = j \mid X(n-1) = i\big) = \sum_{\forall j} \frac{P\big(X(n-1) = i \mid X(n) = j\big)\,P\big(X(n) = j\big)}{P\big(X(n-1) = i\big)} = \frac{P\big(X(n-1) = i\big)}{P\big(X(n-1) = i\big)} = 1 \tag{3.6}$$

- $\mathbf{P}$ is a **stochastic matrix**, i.e. a matrix which rows sum 1.

Example

- Assume a terminal can be in **3 states**:
  - State 1: Idle.
  - State 2: Active without sending data.
  - State 3: Active and sending data at a rate v bps.



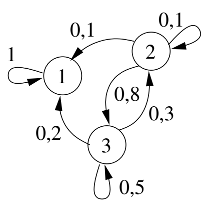
$$\mathbf{P} = \begin{array}{c} \textbf{to state} \\ \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0{,}8 & 0{,}2 & 0 \\ 0 & 0{,}2 & 0{,}8 \\ 0{,}2 & 0{,}3 & 0{,}5 \end{bmatrix} \end{array} \begin{array}{l} \\ \\ 1 \quad \textbf{from} \\ 2 \quad \textbf{state} \\ 3 \end{array} \tag{3.7}$$

- The **average transmission rate** (throughput), v_a, is:

$$v_a = P\big(\text{the terminal is in state } 3\big) \times v \tag{3.8}$$

3.4 Absorbing Chains

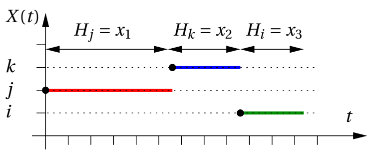
- It is possible to have chains with **absorbing states**.
- A state i is absorbing if $p_{ii} = 1$.
- Example: State 1 is absorbing.



$$\mathbf{P} = \begin{array}{c} \textbf{to state} \\ \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0{,}1 & 0{,}1 & 0{,}8 \\ 0{,}2 & 0{,}3 & 0{,}5 \end{bmatrix} \end{array} \begin{array}{l} \\ \\ 1 \quad \textbf{from} \\ 2 \quad \textbf{state} \\ 3 \end{array} \tag{3.9}$$

3.5 Sojourn or Holding Time

- **Sojourn** or **holding time** in state k: Is the RV H_k equal to the number of steps that the chain remains in state k before leaving to a different state:



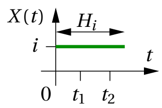
- The Markov property implies:

$$H_i(n) = P\left(H_i = n\right) = p_{ii}^{n-1}\,(1-p_{ii}),\, n \ge 1 \tag{3.10}$$

- Which is a **geometric** distribution with mean:

$$\mathrm{E}[H_i] = \sum_{n=1}^{\infty} n\,P\left(H_i = n\right) = \frac{1}{1-p_{ii}}. \tag{3.11}$$

The geometric distribution satisfies the Markov property



Proof.

- Markov property:
$P\left(X(n_2) = i \mid X(n_1) = i, X(n_0) = i\right) = P\left(X(n_2) = i \mid X(n_1) = i\right)$
- Thus, the Markov property **in terms of the sojourn time** can be written as:

$$P\left(H_i > n_2 - n_0 \mid H_i > n_1 - n_0\right) = P\left(H_i > n_2 - n_1\right) \tag{3.12}$$

- Since

$$P\left(H_i > k\right) = 1 - P\left(H_i \le k\right) = 1 - \sum_{n=1}^{k} p^{n-1}(1-p) = 1-(1-p)\frac{1-p^k}{1-p} = p^k \tag{3.13}$$

- We have:

$$P\left(H_i > n_2 - n_0 \mid H_i > n_1 - n_0\right) = \frac{P\left(H_i > n_2 - n_0, H_i > n_1 - n_0\right)}{P\left(H_i > n_1 - n_0\right)} = \frac{P\left(H_i > n_2 - n_0\right)}{P\left(H_i > n_1 - n_0\right)} = \frac{p^{n_2 - n_0}}{p^{n_1 - n_0}} = p^{n_2 - n_1} = P\left(H_i > n_2 - n_1\right) \tag{3.14}$$

□

3.6 n-step transition probabilities

- Transition probabilities: $p_{ij} = P\left(X(n) = j \mid X(n-1) = i\right)$
- In matrix form:

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots \\ p_{21} & p_{22} & \cdots \\ \cdots & \cdots & \cdots \end{bmatrix} \tag{3.15}$$

- We define the **n-step** transition probabilities:

$$p_{ij}(n) = P\left(X(n) = j \mid X(0) = i\right) \tag{3.16}$$

$$\mathbf{P}(n) = \begin{bmatrix} p_{11}(n) & p_{12}(n) & \cdots \\ p_{21}(n) & p_{22}(n) & \cdots \\ \cdots & \cdots & \cdots \end{bmatrix} \tag{3.17}$$

- $\mathbf{P}$ and $\mathbf{P}(n)$ are **stochastic matrices**: Their rows sum 1.

3.7 State Probabilities

- Define the probability of being in state i at step n:

$$\pi_i(n) = P\left(X(n) = i\right) \tag{3.18}$$

- In vector form (row vector)

$$\boldsymbol{\pi}(n) = (\pi_1(n), \pi_2(n), \cdots) = (P\left(X(n) = 1\right), P\left(X(n) = 2\right), \cdots). \tag{3.19}$$

- Thus, the vector $\boldsymbol{\pi}(n)$ is the distribution of the random variable $X(n)$, and it is called the **state probability at step** n.
- Law of total prob. $P(A) = \sum_n P(A \cap B_n) = \sum_n P(A|B_n)P(B_n)$:

$$\begin{aligned} \pi_i(n) &= \sum_k P\left(X(n-1) = k\right) P\left(X(n) = i \mid X(n-1) = k\right) = \sum_k \pi_k(n-1)\,p_{ki} \\ \pi_i(n) &= \sum_k P\left(X(0) = k\right) P\left(X(n) = i \mid X(0) = k\right) = \sum_k \pi_k(0)\,p_{ki}(n) \end{aligned} \tag{3.20}$$

- In matrix form:

$$\begin{aligned} \boldsymbol{\pi}(n) &= \boldsymbol{\pi}(n-1)\,\mathbf{P} \\ \boldsymbol{\pi}(n) &= \boldsymbol{\pi}(0)\,\mathbf{P}(n) \end{aligned} \tag{3.21}$$

where $\boldsymbol{\pi}(0)$ is the **initial distribution**.

- Iterating

$$\boldsymbol{\pi}(n) = \boldsymbol{\pi}(n-1)\,\mathbf{P} = \boldsymbol{\pi}(n-2)\,\mathbf{P}\,\mathbf{P} = \boldsymbol{\pi}(n-3)\,\mathbf{P}\,\mathbf{P}\,\mathbf{P} = \cdots = \boldsymbol{\pi}(0)\,\mathbf{P}^n \tag{3.22}$$

- Thus:

$$\boldsymbol{\pi}(n) = \boldsymbol{\pi}(0)\,\mathbf{P}(n) = \boldsymbol{\pi}(0)\,\mathbf{P}^n \tag{3.23}$$

3.8 Chapman-Kolmogorov Equations

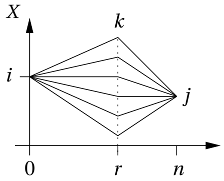
$$p_{ij}(n) = \sum_k p_{ik}(r)\; p_{kj}(n-r) \tag{3.24}$$

Proof.

$$\begin{aligned} p_{ij}(n) &= P\big(X(n) = j \mid X(0) = i\big) = \\ &\sum_k P\big(X(n) = j, X(r) = k \mid X(0) = i\big) \\ &= \sum_k \frac{P\big(X(n) = j, X(r) = k, X(0) = i\big)}{P\big(X(0) = i\big)} \times \frac{P\big(X(r) = k, X(0) = i\big)}{P\big(X(r) = k, X(0) = i\big)} = \\ &\sum_k P\big(X(n) = j \mid X(r) = k, X(0) = i\big) \times P\big(X(r) = k \mid X(0) = i\big) = \\ &\sum_k P\big(X(n) = j \mid X(r) = k\big)\, P\big(X(r) = k \mid X(0) = i\big) \\ &= \sum_k p_{ik}(r)\, p_{kj}(n-r) \end{aligned}$$

□

- Graphical interpretation:



- In matrix form:

$$\boxed{\mathbf{P}(n) = \mathbf{P}(r)\,\mathbf{P}(n-r) \qquad (3.25)}$$

- Particularly:

$$\mathbf{P}(n) = \mathbf{P}(1)\,\mathbf{P}(n-1) = \mathbf{P}\,\mathbf{P}(n-1) = \mathbf{P}(n-1)\,\mathbf{P} \qquad (3.26)$$

- Iterating:

$$\boxed{\mathbf{P}(n) = \mathbf{P}^n \qquad (3.27)}$$

- Thus:

$$\boxed{\boldsymbol{\pi}(n) = \boldsymbol{\pi}(0)\,\mathbf{P}(n) = \boldsymbol{\pi}(0)\,\mathbf{P}^n \qquad (3.28)}$$

Chapter 4
Transient Solution

4.1 Close Form Solution

- If we are interested in the **transient evolution** we shall study $\boldsymbol{\pi}(n) = \boldsymbol{\pi}(0)\,\mathbf{P}^n$.
- If we can **diagonalize P**, we can obtain the transient evolution in **close form**.
- **P** can be **diagonalized** if **P** can be decomposed as:

$$\mathbf{P} = \mathbf{L}^{-1}\,\Lambda\,\mathbf{L} \qquad (4.1)$$

where **L** is some invertible matrix and Λ is the diagonal matrix

$$\Lambda = \mathrm{diag}(\lambda_1, \cdots \lambda_N) = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_N \end{bmatrix} \qquad (4.2)$$

with λ_l, $l = 1, \cdots N$ the **eigenvalues** of **P**.

- Assume a **finite DTMC** with N states. Then $\mathbf{P} = \mathbf{P}^{N\times N}$.
- Assume that **P** can be **diagonalized**: $\mathbf{P} = \mathbf{L}^{-1}\,\Lambda\,\mathbf{L}$, where Λ is the diagonal matrix $\Lambda = \mathrm{diag}(\lambda_1, \cdots \lambda_N)$, with λ_l, $l = 1, \cdots N$ the eigenvalues of **P**.
- But $\mathbf{L}^{-1}\,\mathrm{diag}(\lambda_1^n, \cdots \lambda_N^n)\,\mathbf{L}$ are linear combinations of $\lambda_1^n, \cdots \lambda_N^n$. Thus, the probability of being in state i is given by:

$$\boxed{\pi_i(n) = (\boldsymbol{\pi}(n))_i = \sum_{l=1}^{N} a_i^{(l)}\,\lambda_l^n \qquad (4.3)}$$

where the **unknown coefficients** $a_i^{(l)}$ can be obtained solving the system of equations:

$$\sum_{l=1}^{N} a_i^{(l)}\,\lambda_l^n = (\boldsymbol{\pi}(n))_i = (\boldsymbol{\pi}(0)\,\mathbf{P}^n)_i,\; n = 0, \cdots N-1 \qquad (4.4)$$

4.2 Eigenvalues

- The **eigenvalues** λ_l of a matrix **A** are scalars that satisfy: $\boldsymbol{l}\,\mathbf{A} = \lambda_l\,\boldsymbol{l}$ (or $\mathbf{A}\,\boldsymbol{r} = \lambda_l\,\boldsymbol{r}$) for some row vectors $\boldsymbol{l}$ (column vectors $\boldsymbol{r}$), referred to as *left* **and** *right* **eigenvectors**, respectively.

$$\begin{aligned} \boldsymbol{l}\,\mathbf{A} = \lambda_l\,\boldsymbol{l} &\Rightarrow \boldsymbol{l}\,(\mathbf{A} - \mathbf{I}\lambda_l) = 0 \Rightarrow \det(\lambda_l\,\mathbf{I} - \mathbf{A}) = 0 \\ \mathbf{A}\,\boldsymbol{r} = \lambda_l\,\boldsymbol{r} &\Rightarrow (\mathbf{A} - \mathbf{I}\lambda_l)\boldsymbol{r} = 0 \Rightarrow \det(\lambda_l\,\mathbf{I} - \mathbf{A}) = 0 \end{aligned} \qquad (4.5)$$

- Thus, λ_l solve the **characteristic polynomial** $\det(\lambda\,\mathbf{I} - \mathbf{A}) = 0$.
- Note that, in general, *left* **and** *right* **eigenvectors** are different, but eigenvalues are the same (they solve the same **characteristic polynomial**).
- A matrix can be **diagonalized** if all eigenvalues are single (multiplicity = 1). If a matrix cannot be diagonalized it is called **defective**.

4.3 Determinant

$$\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}\,a_{22} - a_{12}\,a_{21} \qquad (4.6)$$

$$\det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{matrix} +a_{11}\,a_{22}\,a_{33} + a_{12}\,a_{23}\,a_{31} + a_{21}\,a_{32}\,a_{13} \\ -a_{31}\,a_{22}\,a_{13} - a_{12}\,a_{21}\,a_{33} - a_{23}\,a_{32}\,a_{11} \end{matrix} \qquad (4.7)$$

- **Cofactor Formula**: expanding along a row i:

$$\det \mathbf{A} = \sum_{j=1}^{N} a_{ij}\,(-1)^{i+j} \det M_{ij}, \qquad (4.8)$$

where the **minor matrices** M_{ij} are obtained removing the row i and column j from **A**. $(-1)^{i+j}\det M_{ij}$ is called the **cofactor** of a_{ij}.

$$\det \mathbf{A} = \prod \text{eigenvalues of } \mathbf{A} \qquad (4.9)$$

$$\text{trace } \mathbf{A} = \sum \text{eigenvalues of } \mathbf{A} \qquad (4.10)$$

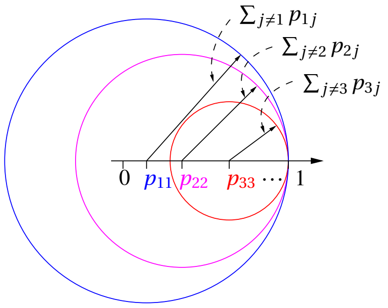
where trace $\mathbf{A} = \sum$ elements of the diagonal of **A**.

4.4 Eigenvalues of a Stochastic Matrix

- **P** has **an eigenvalue equal to 1** ($\mathbf{P}\,\boldsymbol{x} = \lambda\,\boldsymbol{x}$, for $\lambda = 1$).

  Proof: $\mathbf{P}\,\mathbf{e} = \mathbf{e}$, where $\mathbf{e} = \begin{bmatrix} 1 & 1 & \cdots \end{bmatrix}^{\mathrm{T}}$ is a column vector of 1 (all rows of **P** add to 1).
- All eigenvalues of **P** are $|\lambda_l| \le 1$.

Proof. Using Gerschgorin's theorem *The eigenvalues of a matrix* $\mathbf{P}_{n\times n}$ *lie within the union of the n circular disks with center* p_{ii} *and radius* $\sum_{j\neq i} |p_{ij}|$ in $\mathbb{C}$. Since $\sum_j p_{ij} = 1$, the property is proved.



Gershgorin discs
of a stochastic matrix

■

Proof. of Gerschgorin's theorem
From $\mathbf{P}\boldsymbol{x} = \lambda\,\boldsymbol{x}$ we have

$$\sum_j p_{ij}\,x_j = \lambda\,x_i \qquad \forall i \in \{1,\dots,n\}. \tag{4.11}$$

We choose i such that $|x_i| = \max_j |x_j|$. Thus, $\sum_{j\neq i} p_{ij}\,x_j = \lambda\,x_i - p_{ii}\,x_i$, and

$$|\lambda - p_{ii}| = \left|\sum_{j\neq i} p_{ij}\,\frac{x_j}{x_i}\right| \le \sum_{j\neq i}\left|p_{ij}\,\frac{x_j}{x_i}\right| \le \sum_{j\neq i}|p_{ij}| \tag{4.12}$$

and the equation $|\mathbf{x}-\mathbf{c}| \le \mathbf{r}$, $x,c \in \mathbb{C}, r \in \mathbb{R}$ is a disk of center c and radius r in $\mathbb{C}$. ■

Example

- Assume a DTMC with

$$\mathbf{P} = \begin{bmatrix} 4/5 & 1/5 \\ 2/5 & 3/5 \end{bmatrix}$$

- We want the probability of being in state 2 in n steps starting from state 1: $\pi_2(n)$ with $\boldsymbol{\pi}(0) = \begin{bmatrix}1 & 0\end{bmatrix}$.

Solution

- It can be easily found that the **eigenvalues** of $\mathbf{P}$ are $\lambda_1 = 1$ and $\lambda_2 = 2/5$.

$$\pi_2(n) = \lambda_1^n a + b\,\lambda_2^n = a + b\,(2/5)^n \tag{4.13}$$

- Imposing the **boundary conditions** $\pi_i(n) = (\boldsymbol{\pi}(0)\,\mathbf{P}^n)_i$:

$$\begin{aligned} \pi_2(0) &= a + b = (\begin{bmatrix}1 & 0\end{bmatrix}\mathbf{P}^0)_2 = (\mathbf{P}^0)_{12} = 0 \\ \pi_2(1) &= a + b\,(2/5) = (\begin{bmatrix}1 & 0\end{bmatrix}\mathbf{P}^1)_2 = (\mathbf{P})_{12} = 1/5 \end{aligned} \tag{4.14}$$

we have that $a = 1/3$, $b = -1/3$, thus:

$$\boxed{\begin{aligned} \pi_2(n) &= 1/3 - 1/3\,(2/5)^n, \quad n \ge 0 \\ \pi_1(n) &= 1 - \pi_2(n) = 2/3 + 1/3\,(2/5)^n, \quad n \ge 0 \end{aligned}} \tag{4.15}$$

4.5 Chain with a Defective Matrix

- What if $\mathbf{P}$ cannot be diagonalized? (**defective** matrix).
- Let λ_l, $l = 1,\cdots L$ be the eigenvalues of $\mathbf{P}^{N\times N}$, each with multiplicity k_l ($k_l \ge 1$, $\sum_l k_l = N$), and a possible eigenvalue $\lambda_1 = 0$ with multiplicity k_1. Then [1]:

$$\boxed{\begin{aligned} \pi_j(n) = \sum_{m=0}^{k_1-1} a_j^{(1,m)}\,I(n=m) + \sum_{l=2}^{L}\lambda_l^n \sum_{m=0}^{k_l-1} a_j^{(l,m)}\,n^m, \\ 1 \le j \le N, n \ge 0 \end{aligned}} \tag{4.16}$$

$I(n=m)$ is the indicator func.: $I(n) = 1$ if $n = m$, $I(n) = 0$ if $n \ne m$.

Example

- Assume a DTMC with

$$\mathbf{P} = \begin{bmatrix} 3/4 & 1/4 & 0 \\ 0 & 3/4 & 1/4 \\ 1 & 0 & 0 \end{bmatrix} \tag{4.17}$$

- We want the probability of being in state 1 in n steps starting from state 1: $\pi_1(n)$ with $\pi_1(0) = 1$.
- It can be easily found that the **eigenvalues** of $\mathbf{P}$ are $\lambda_1 = 1$ and $\lambda_2 = 1/4$ with multiplicity 2. We guess:

$$\pi_1(n) = a + 1/4^n(b + c\,n) \tag{4.18}$$

- Imposing $\pi_1(0) = 1$, $\pi_1(1) = 3/4$, $\pi_1(2) = (3/4)^2$, we have:

$$\boxed{\pi_1(n) = \frac{4}{9} + \frac{1}{4^n}\left(\frac{5}{9} + \frac{2}{3}\,n\right)} \tag{4.19}$$

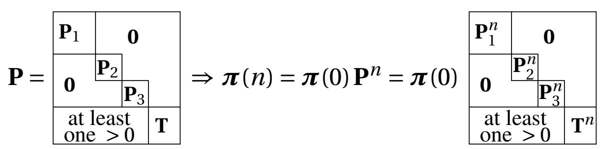
Chapter 5
Classification of States

Objective

- Identify the different **types of behavior** that the chain can have.
- Introduce the concepts of **first passage probability** and **mean recurrence time**.

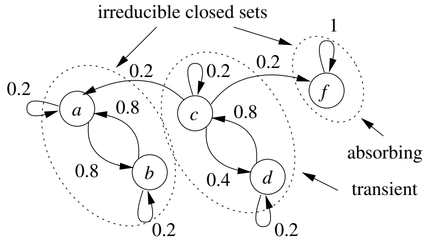
5.1 Irreducibility

- A state j is said to **communicate** with i, $i \leftrightarrow j$, if $p_{ij}(m_1) > 0$, $p_{ji}(m_2) > 0$ for some $m_1, m_2 \ge 0$.
- We define an **irreducible closed set, ICS** C_k as a set where all states communicate with each other, and have no transitions to other states out of the set:
$i \leftrightarrow j, \forall i, j \in C_k$ and $p_{ij} = 0, \forall i \in C_k, j \notin C_k$
(note that for $i \in C_k, j \notin C_k$ we have: $p_{ij}(2) = \sum_k p_{ik}\,p_{kj} = 0$, since $p_{ik} = 0$ if $k \notin C_k$, and $p_{kj} = 0$ if $k \in C_k$. Thus, $p_{ij}(n) = 0, \forall n$.)
- An **absorbing state** form an ICS of only one element. This state, i, must have $p_{ii} = 1$, $p_{ij} = 0\;\forall j \ne i$.
- **Transient states** do not belong to any ICS.
- A MC is **irreducible** if all the states form a unique ICS.
- Assume a MC has M **ICSs**: By properly numbering the states, we can write $\mathbf{P}$ as an M block diagonal matrix with the probabilities of the transient states in the last rows.
- **Example**, if $M = 3$:



- Note that **the M sub-matrices are stochastic** (their rows sum 1).

Example



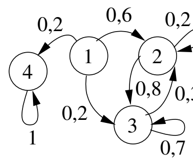
$$\mathbf{P} = \begin{array}{c|c} & \begin{array}{ccccc} a & b & f & c & d \end{array} \\ \begin{array}{c} a \\ b \\ f \\ c \\ d \end{array} & \left[\begin{array}{cc|c|cc} 0{,}2 & 0{,}8 & 0 & 0 & 0 \\ 0{,}8 & 0{,}2 & 0 & 0 & 0 \\ \hline 0 & 0 & 1{,}0 & 0 & 0 \\ \hline 0{,}2 & 0 & 0{,}2 & 0{,}2 & 0{,}4 \\ 0 & 0 & 0 & 0{,}8 & 0{,}2 \end{array}\right] \end{array}$$

$$\mathbf{P}^{\infty} = \begin{array}{c|c} & \begin{array}{ccccc} a & b & f & c & d \end{array} \\ \begin{array}{c} a \\ b \\ f \\ c \\ d \end{array} & \left[\begin{array}{cc|c|cc} 0{,}5 & 0{,}5 & 0 & 0 & 0 \\ 0{,}5 & 0{,}5 & 0 & 0 & 0 \\ \hline 0 & 0 & 1{,}0 & 0 & 0 \\ \hline 0{,}25 & 0{,}25 & 0{,}5 & 0 & 0 \\ 0{,}25 & 0{,}25 & 0{,}5 & 0 & 0 \end{array}\right] \end{array}$$

- What is the meaning of the probabilities in $\mathbf{P}^{\infty}$? (recall that $(\mathbf{P}^n)_{ij} = p_{ij}(n) = P\big(X(n) = j \mid X(0) = i\big)$).

5.2 Transient and Recurrent

- **Recurrent**: States that, being visited, they are visited again with probability 1. They are visited an infinite number of times when $n \to \infty$.
- **Transient**: States that, being visited, have a probability > 0 of never being visited again. They are visited a finite number of times when $n \to \infty$.
- **Absorbing**: A single (recurrent) state where the chain remains with probability $= 1$.

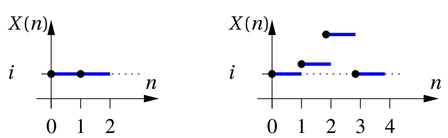


State 1 is transient
States 2 and 3 are recurrent
State 4 is absorbing

5.3 First Passage (Transition) Probabilities

- To derive a classification criteria, we shall study the distribution of the number of steps to **go for the first time from a state i another state j**. Definition:

$$f_{ii}(n) = P\begin{pmatrix}\text{first transition into state } i \\ \text{in } n \text{ steps starting from } i\end{pmatrix} \tag{5.1}$$



first transition in 1 step
first transition in 3 steps

- Do **not confuse** with the n-step transition probability $p_{ii}(n)$, where the state i can be visited in the intermediate states.

Relation between $f_{ii}(n)$ and $p_{ii}(n)$

- $f_{ii}(n)$ and $p_{ii}(n)$ satisfy:

$$\begin{aligned} f_{ii}(1) &= p_{ii}(1) \\ p_{ii}(n) &= \sum_{l=1}^{n} f_{ii}(l)\, p_{ii}(n-l),\ n >= 1 \end{aligned} \tag{5.2}$$

- The probability that the MC **eventually enters state i starting from i** is given by:

$$f_{ii} = \sum_{n=1}^{\infty} f_{ii}(n) \tag{5.3}$$

- If $f_{ii} = 1$ we say i is a **recurrent state**.
- If $f_{ii} < 1$ we say i is a **transient state**.

5.4 Mean Recurrence Time

- When $f_{ii} = 1$, we define the **mean recurrence time** m_{ii} as:

$$m_{ii} = \sum_{n=1}^{\infty} n\, f_{ii}(n) \tag{5.4}$$

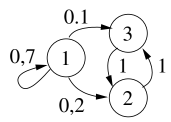
- m_{ii} is the **average number of steps to eventually reach i starting from i**. If $f_{ii} < 1$ (**transient state**) then we define $m_{ii} = \infty$.
- Classification of **recurrent states** ($f_{ii} = 1$):
  - If $m_{ii} = \infty$ the state is **null recurrent**: it takes an ∞ time to reach the state after leave it. Can only happen in chains with an infinite number of states.
  - If $m_{ii} < \infty$ the state is **positive recurrent**: the state is reached in a finite time after leave it.

5.5 Property of States

In **finite MC**:

1. States can be only of type positive recurrent or transient.
2. At least one state must be positive recurrent.
3. There are not null recurrent states.

- **Example**:



- State 1 is transient. States 2 and 3 are positive recurrent.

Generalization to Any State Pair

- Analogously to $f_{ii}(n)$, we define the probability of the **first passage to state j starting from any state i** in n steps: $f_{ij}(n)$.
- $f_{ij}(n)$ and $p_{ij}(n)$ satisfy:

$$p_{ij}(n) = \sum_{l=1}^{n} f_{ij}(l)\, p_{ij}(n-l),\; n \geq 1 \tag{5.5}$$

- **When** $f_{ij} = 1$, the average number of steps to eventually reach j starting from i, m_{ij} is given by:

$$\boxed{m_{ij} = \sum_{n=1}^{\infty} n\, f_{ij}(n)} \tag{5.6}$$

- If state j can not be reached starting from state i with probability one (**if** $f_{ij} < 1$), then we define $m_{ij} = \infty$.

5.6 Recursive Equation for the First Passage Probabilities

- Recall that the The probability that the MC **eventually enters state j starting from** i is given by: $f_{ij} = \sum_{n=1}^{\infty} f_{ij}(n)$
- f_{ij} can be computed as follows: Assume we are in i. With probability p_{ij} we will go to j in one step. Otherwise, we will go to k, $k \neq j$, and then we will reach j with probability f_{kj}. Thus:

$$\boxed{f_{ij} = p_{ij} + \sum_{k \neq j} p_{ik} f_{kj}} \tag{5.7}$$

- If there are more than 1 **absorbing states**, we can compute the probability to reach them using this method (if there is only 1, say j, then $f_{ij} = 1, \forall i$).

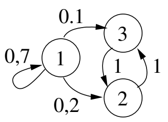
5.7 Recursive Equation for the Mean Recurrence Time

- Recall that the **mean recurrence time** $m_{ij} = \sum_{n\geq 1} n\, f_{ij}(n)$ is the average number of steps to eventually reach j starting from i, i.e. it is the mean first passage time from state i to j.
- **When** $f_{ij} = 1$, m_{ij} can be computed as follows: Assume we are in i. With probability p_{ij} we will go to j in one step. Otherwise, we will go to k, $k \neq j$, and then it will take m_{kj} steps to reach j. Thus:

$$\boxed{m_{ij} = p_{ij} + \sum_{k \neq j} p_{ik}(1 + m_{kj}) = 1 + \sum_{k \neq j} p_{ik}\, m_{kj}} \tag{5.8}$$

since $\sum_j p_{ij} = 1$.

Example: Recurrence Times Using the Definition



$$\mathbf{P} = \begin{bmatrix} 0{,}7 & 0{,}2 & 0{,}1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$f_{21}(n) = f_{31}(n) = 0$$
$$f_{11}(n) = 0{,}7\, I(n=1)$$
$$f_{22}(n) = f_{33}(n) = I(n=2)$$
$$f_{23}(n) = f_{32}(n) = I(n=1)$$

$$f_{12}(n) = \begin{cases} 0{,}2, & n=1 \\ 0{,}7^{n-1}\, 0{,}2 + 0{,}7^{n-2}\, 0{,}1, & n>1 \end{cases}$$

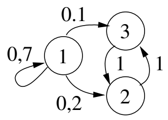
$$f_{13}(n) = \begin{cases} 0{,}1, & n=1 \\ 0{,}7^{n-1}\, 0{,}1 + 0{,}7^{n-2}\, 0{,}2, & n>1 \end{cases}$$

$$f_{11} = 0{,}7$$
$$f_{12} = f_{13} = 1 \qquad f_{22} = f_{23} = 1$$
$$f_{32} = f_{33} = 1 \qquad f_{21} = f_{31} = 0$$

$$\mathbf{M} = (m_{ij}) = \begin{bmatrix} \infty & 11/3 & 12/3 \\ \infty & 2 & 1 \\ \infty & 1 & 2 \end{bmatrix}$$

- State 1 is **transient**. States 2 and 3 are **recurrent**.

Example: First Passage Probability Using Recursion



$$\mathbf{P} = \begin{bmatrix} 0{,}7 & 0{,}2 & 0{,}1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

- We have:

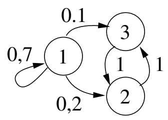
$$f_{12} = p_{11} f_{12} + p_{12} + p_{13} f_{32} \tag{5.9}$$

- Clearly $f_{32} = 1$, thus:

$$f_{12} = 0{,}7 f_{12} + 0{,}2 + 0{,}1 \times 1 \Rightarrow \mathbf{f_{12} = 1} \tag{5.10}$$

as before.

Example: Mean Recurrence Time Using Recursion



$$\mathbf{P} = \begin{bmatrix} 0{,}7 & 0{,}2 & 0{,}1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

- We have:

$$m_{12} = p_{12} + p_{11}(1 + m_{12}) + p_{13}(1 + m_{32}) = 1 + p_{11}\, m_{12} + p_{13}\, m_{32} \tag{5.11}$$

- Clearly $m_{32} = 1$, thus:

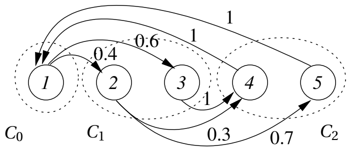
$$m_{12} = 1 + 0{,}7\, m_{12} + 0{,}1 \times 1 \Rightarrow \mathbf{m_{12} = 11/3}. \tag{5.12}$$

5.8 Periodic states

- A recurrent state j is **periodic** with period $d > 1$ if j can only be reached after leaving it with a multiple of d steps.
- If $d = 1$ the state is aperiodic.
- Any **periodic irreducible chain** can be partitioned in d **cyclic classes** $C_0, \cdots C_{d-1}$ such that at each step a transition occur from class C_i to $C_{(i+1) \bmod d}$.
- By properly numerating the states, the transition matrix can be written as (the sub-matrices $\mathbf{A}_i$ may not be square):

$$\mathbf{P} = \begin{array}{c} \\ C_0 \\ C_1 \\ \cdots \\ C_{d-1} \end{array} \begin{array}{c} \begin{array}{ccccc} C_0 & C_1 & C_2 & \cdots & C_{d-1} \end{array} \\ \begin{bmatrix} 0 & \mathbf{A}_1 & 0 & \cdots & 0 \\ 0 & 0 & \mathbf{A}_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathbf{A}_{d-1} & 0 & 0 & \cdots & 0 \end{bmatrix} \end{array} \tag{5.13}$$

Example



$$\mathbf{P} = \begin{bmatrix} 0 & 0.4 & 0.6 & 0 & 0 \\ 0 & 0 & 0 & 0.3 & 0.7 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{P}^2 = \begin{bmatrix} 0 & 0 & 0 & 0.72 & 0.28 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0.4 & 0.6 & 0 & 0 \\ 0 & 0.4 & 0.6 & 0 & 0 \end{bmatrix},$$

$$\mathbf{P}^3 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0.4 & 0.6 & 0 & 0 \\ 0 & 0.4 & 0.6 & 0 & 0 \\ 0 & 0 & 0 & 0.72 & 0.28 \\ 0 & 0 & 0 & 0.72 & 0.28 \end{bmatrix},$$

$$\mathbf{P}^4 = \begin{bmatrix} 0 & 0.4 & 0.6 & 0 & 0 \\ 0 & 0 & 0 & 0.72 & 0.28 \\ 0 & 0 & 0 & 0.72 & 0.28 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}, \cdots$$

- In periodic chains $\mathbf{P}^n$ **does not converge**.

Chapter 6
Steady State

6.1 Limiting Distribution

- Probability of being in state i at step n:

$$\pi_i(n) = P\big(X(n) = i\big). \tag{6.1}$$

In vector form (row vector)

$$\boldsymbol{\pi}(n) = (\pi_1(n), \pi_2(n), \cdots). \tag{6.2}$$

- The evolution of the chain depends on the initial distribution $\boldsymbol{\pi}(0)$.
- If we are interested in the **transient** evolution we shall study

$$\boldsymbol{\pi}(n) = \boldsymbol{\pi}(0)\,\mathbf{P}^n. \tag{6.3}$$

- If we are interested the **steady state** we shall be interested in the **limiting distribution** (if the limit exists):

$$\boldsymbol{\pi}(\infty) = (\pi_1(\infty), \pi_2(\infty), \cdots) \tag{6.4}$$

Assume an **irreducible** chain with **positive recurrent** states.

- With infinite steps, we look for a probability converging to a value that depends only on the final state:

$$\pi_j(\infty) = \sum_i \pi_i(0) \lim_{n\to\infty} p_{ij}(n), \forall j \text{ and for any } \boldsymbol{\pi}(0), \tag{6.5}$$

which implies:

$$\pi_j(\infty) = \lim_{n\to\infty} p_{ij}(n) \sum_i \pi_i(0) = p_{ij}(\infty), \forall j \Rightarrow$$

$$\mathbf{P}(\infty) = \lim_{n\to\infty} \mathbf{P}^n = \begin{bmatrix} \boldsymbol{\pi}(\infty) \\ \cdots \\ \boldsymbol{\pi}(\infty) \end{bmatrix} \tag{6.6}$$

- If this limit exists, we call $\mathbf{P}(\infty)$ the **limiting matrix**, and $\boldsymbol{\pi}(\infty)$ the **limiting distribution**.

Example

$$\mathbf{P} = \begin{bmatrix} 0.8 & 0.15 & 0.05 \\ 0.7 & 0.2 & 0.1 \\ 0.5 & 0.3 & 0.2 \end{bmatrix}$$

$$\mathbf{P}^2 = \begin{bmatrix} 0.764 & 0.168 & 0.068 \\ 0.760 & 0.170 & 0.070 \\ 0.752 & 0.174 & 0.074 \end{bmatrix}$$

$$\mathbf{P}^4 = \begin{bmatrix} 0{,}7628 & 0{,}1686 & 0{,}0686 \\ 0{,}7620 & 0{,}1690 & 0{,}0690 \\ 0{,}7604 & 0{,}1698 & 0{,}0698 \end{bmatrix}$$

$$\mathbf{P}^8 = \begin{bmatrix} 0.762500 & 0.168750 & 0.068750 \\ 0.762499 & 0.168750 & 0.068750 \\ 0.762497 & 0.168752 & 0.068752 \end{bmatrix}$$

$$\cdots$$

$$\Rightarrow \boldsymbol{\pi}(\infty) = (0.76250, 0.16875, 0.06875)$$

6.2 Stationary distribution

- We have:

$$\pi_i(n) = P\big(X(n) = i\big) = \sum_k P\big(X(n-1) = k\big)\, P\big(X(n) = i \mid X(n-1) = k\big) = \sum_k \pi_k(n-1)\, p_{ki} \tag{6.7}$$

- In matrix form: $\boldsymbol{\pi}(n) = \boldsymbol{\pi}(n-1)\,\mathbf{P}$
- If $\pi_i(n) = \pi_i(n-1) = \pi_i\ \forall i$, we call π_i the **stationary probability of state** i, and $\boldsymbol{\pi} = (\pi_1, \pi_2, \cdots)$, the **stationary distribution** of the chain.
- In matrix form (**Global balance equations**):

$$\boxed{\begin{aligned} &\boldsymbol{\pi} = \boldsymbol{\pi}\,\mathbf{P} \\ &\boldsymbol{\pi}\,\mathbf{e} = 1, \mathbf{e} = \begin{bmatrix} 1 & 1 & \cdots \end{bmatrix}^{\mathrm{T}} \end{aligned}} \tag{6.8}$$

- Thus, the stationary distribution is the left-hand eigenvector corresponding to the unit eigenvalue of $\mathbf{P}$.
- $\boldsymbol{\pi}(n) = \boldsymbol{\pi} \Rightarrow \boldsymbol{\pi}(n+1) = \boldsymbol{\pi}(n)\,\mathbf{P} = \boldsymbol{\pi}\,\mathbf{P} = \boldsymbol{\pi} \Rightarrow \boxed{\boldsymbol{\pi}(k) = \boldsymbol{\pi}, k \geq n}$
- Do not confuse the **limiting distribution** $\boldsymbol{\pi}(\infty)$ and the **stationary distribution** $\boldsymbol{\pi} = \boldsymbol{\pi}\,\mathbf{P}$.
- $\boldsymbol{\pi}(\infty)$ and $\boldsymbol{\pi}$ **may not be the same**, e.g. in **periodic chains** $\boldsymbol{\pi}(\infty)$ does not exists ($\mathbf{P}$ does not converge), but we can compute the stationary distribution.

- **Example:** the periodic chain

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \tag{6.9}$$

has the stationary distribution

$$\boldsymbol{\pi} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix}. \tag{6.10}$$

6.3 Numerical Solution

Replace one equation method

$$\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$$
$$\boldsymbol{\pi}\mathbf{e} = 1, \mathbf{e} = \begin{bmatrix} 1 & 1 & \cdots \end{bmatrix}^{\mathrm{T}}$$

We solve the equation $\boldsymbol{\pi}(\mathbf{I}-\mathbf{P}) = 0$ replacing the last equation by $\boldsymbol{\pi}\mathbf{e} = 1$:

$$\boldsymbol{\pi} \begin{bmatrix} p_{11}-1 & p_{12} & \cdots p_{1n-1} & 1 \\ p_{21} & p_{22}-1 & \cdots p_{2n-1} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ p_{n1} & p_{n2} & \cdots p_{nn-1} & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \tag{6.11}$$

Examples

- **Replace one equation method**: $\mathbf{P} = \begin{bmatrix} 0.8 & 0.15 & 0.05 \\ 0.7 & 0.2 & 0.1 \\ 0.5 & 0.3 & 0.2 \end{bmatrix}$
- With **octave** (matlab clone):

```
octave:1> P
    =[0.8,0.15,0.05;0.7,0.2,0.1;0.5,0.3,0.2];
octave:2> s=size(P,1); # number of rows.
octave:3> [zeros(1,s-1),1] / ...
> [eye(s,s-1)-P(1:s,1:s-1), ones(s,1)]
ans =
0.762500   0.168750   0.068750
```

- With **R**

```
> P <- matrix(nc=3, byr=T,
+ c(0.8,0.15,0.05,0.7,0.2,0.1,0.5,0.3,0.2))
> s <- nrow(P)
> solve(t(cbind(P[,1:(s-1)]-diag(nr=s,nc=s-1),
+ rep(1,s))), c(rep(0,s-1),1))
[1] 0.76250 0.16875 0.06875
```

NOTE: $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P} \Rightarrow \boldsymbol{\pi}^{\mathrm{T}} = \mathbf{P}^{\mathrm{T}}\boldsymbol{\pi}^{\mathrm{T}}$. The transpose operator in R is `t()`.

6.4 Global balance equations

- Why are they called Global balance equations?

$$\left.\begin{array}{ll} \boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P} \Rightarrow & \pi_j = \sum_{i=0}^{\infty} \pi_i p_{ij} \\ \sum_{i=0}^{\infty} p_{ji} = 1 \Rightarrow & \pi_j \sum_{i=0}^{\infty} p_{ji} = \pi_j \end{array}\right\} \Rightarrow \sum_{i=0}^{\infty} \pi_i p_{ij} = \pi_j \sum_{i=0}^{\infty} p_{ji} \tag{6.12}$$

$$\begin{array}{ll} \sum_{i=0}^{\infty} \pi_i p_{ij} & \Rightarrow \text{Frequency of } \textbf{transitions entering state } j \\ \pi_j \sum_{i=0}^{\infty} p_{ji} & \Rightarrow \text{Frequency of } \textbf{transitions leaving state } j \end{array} \tag{6.13}$$

- In **stationary regime**, the frequency of transitions leaving state j is equal to the frequency of transitions entering state j.

Flux Balancing

- Define the **flux** F_{uv} from state u to v:

$$F_{uv} = \pi_u p_{uv} \tag{6.14}$$

- and the flux from set of states U to V:

$$F(U,V) = \sum_{u\in U}\sum_{v\in V} F_{uv} \tag{6.15}$$

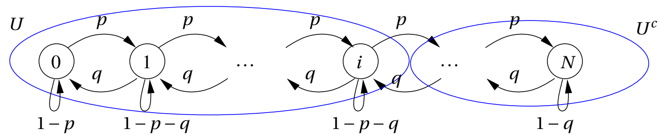
- From the Global balance equations we have:

$$\sum_{i=0}^{\infty} \pi_i p_{ij} = \pi_j \sum_{i=0}^{\infty} p_{ji} \Rightarrow \sum_{i\in U} F_{ij} + \sum_{i\notin U} F_{ij} = \sum_{i\in U} F_{ji} + \sum_{i\notin U} F_{ji} \tag{6.16}$$

- Adding for $j \in U$:

$$\sum_{j\in U}\sum_{i\in U} F_{ij} + \sum_{j\in U}\sum_{i\notin U} F_{ij} = \sum_{j\in U}\sum_{i\in U} F_{ji} + \sum_{j\in U}\sum_{i\notin U} F_{ji} \Rightarrow$$
$$\sum_{j\in U}\sum_{i\notin U} F_{ij} = \sum_{j\in U}\sum_{i\notin U} F_{ji} \Rightarrow \boxed{F(U,U^c) = F(U^c,U)}$$

Solution Using Flux Balancing



- Flux balancing $\Rightarrow$ $p\pi_i = q\pi_{i+1}$
- Iterating: $\pi_1 = \rho\pi_0, \pi_2 = \rho\pi_1 = \rho\rho\pi_0, \cdots, \Rightarrow$

$$\boxed{\pi_i = \rho^i \pi_0, i = 1, \cdots, N} \quad \text{where: } \rho = \frac{p}{q},$$

- Normalizing: $\sum_{i=0}^{N} \pi_i = 1$ $\boxed{\begin{array}{ll} \pi_0 = \dfrac{1-\rho}{1-\rho^{N+1}}, & p \neq q \\ \pi_0 = \dfrac{1}{N+1}, & p = q \end{array}}$

6.5 Ergodic Chains

Ergodic state positive recurrent and aperiodic state.
Ergodic chain if all states are ergodic.
Theorem: All states of an irreducible Markov chain are of the same type: Transient or positive/null recurrent, and aperiodic/periodic with the same period [2, chapter XV].
Consequences:

- **Finite aperiodic and irreducible** chains are **ergodic** (since all states are positive recurrent).
- **Infinite aperiodic and irreducible** chains can be:
  - **Ergodic**: all the states are positive recurrent (stable chains).
  - **Non ergodic**: all states are null recurrent or transient (unstable chains).

6.5.1 Theorems for ergodic chains

- $\boldsymbol{\pi} = \boldsymbol{\pi}(\infty)$

  Proof. For an **aperiodic irreducible** chain with **positive recurrent** states:

$$\begin{cases} \boldsymbol{\pi}(\infty) & = \boldsymbol{\pi}(0)\,\mathbf{P}(\infty) \\ \mathbf{P}(\infty) & = \lim_{n\to\infty} \mathbf{P}^n = \begin{bmatrix} \boldsymbol{\pi}(\infty) \\ \cdots \\ \boldsymbol{\pi}(\infty) \end{bmatrix} \end{cases} \Rightarrow$$

$$\boldsymbol{\pi}(\infty)\,\mathbf{P} = (\boldsymbol{\pi}(0) \lim_{n\to\infty} \mathbf{P}^n)\,\mathbf{P} = \boldsymbol{\pi}(0)\,\mathbf{P}(\infty) = \boldsymbol{\pi}(\infty)$$

$$\Rightarrow \begin{cases} \boldsymbol{\pi}(\infty)\,\mathbf{P} = \boldsymbol{\pi}(\infty) \\ \boldsymbol{\pi}(\infty)\,\mathbf{e} = 1 \end{cases} \quad \boldsymbol{\pi}(\infty) \text{ satisfies the GBE} \Rightarrow \boldsymbol{\pi} = \boldsymbol{\pi}(\infty) \tag{6.17}$$

  □

- In stationary regime (when $\boldsymbol{\pi}(n)\,\mathbf{P} = \boldsymbol{\pi}(n)$), the **mean number of steps the system remains in state** j during k steps is given by

$$\boxed{k\pi_j.} \tag{6.18}$$

  Proof. Assume the chain in stationary regime at time $t = 0$ ($\pi(0)\,\mathbf{P} = \boldsymbol{\pi}(0)$), and let $j(k)$ be the number of visits to j in k steps: $j(k) = \sum_{i=0}^{k-1} I(X(i) = j)$ ($I(A)$ is the indicator function: $I(A) = 1$ if A occurs, $I(A) = 0$ otherwise):

$$\mathrm{E}[j(k)] = \sum_{i=0}^{k-1} \mathrm{E}[I(X(i) = j)] = \sum_{i=0}^{k-1} P(X(i) = j) = k\pi_j \tag{6.19}$$

  □

- In stationary regime the **mean recurrence time** (mean number of steps between two consecutive visits to state j) is given by

$$\boxed{m_{jj} = 1/\pi_j} \tag{6.20}$$

  Proof. Let $j(k)$ be the number of visits to j in k steps:

$$\pi_j = \lim_{k\to\infty} \frac{j(k)}{k} = \lim_{k\to\infty} \frac{1}{k/j(k)} = 1/m_{jj} \tag{6.21}$$

  □

Chapter 7
Reversed Chain

Definition

- Let $X(n)$ be an **ergodic** MC. The chain $X^r(n) = X(-n)$ is referred to as the **time reversal chain** of $X(n)$.
- **Example**, consider a possible sample path of $X(n)$:

$$\cdots (i_0, n_0), (i_1, n_1), (i_2, n_2), \cdots \tag{7.1}$$

  The same path in the time reversal chain $X^r(n)$ would be:

$$\cdots (i_2, -n_2), (i_1, -n_1), (i_0, -n_0), \cdots \tag{7.2}$$

Properties

- Let p_{ij}, p^r_{ij} be the transition probabilities of $X(n)$ respectively $X^r(n)$, and π_i, π^r_i the stationary distributions of $X(n)$ respectively $X^r(n)$, then:

$$\boxed{\pi_i = \pi^r_i} \tag{7.3}$$

- **Proof**: the mean time in each state is the same for both chains.
- However, **in general** $p_{ij} \neq p^r_{ij}$. For example, $X(n)$ may be able to jump from state i to j, but not from j to $i \Rightarrow X^r(n)$ can jump from j to i, but not from i to j.
- But it must be $p_{ii} = p^r_{ii}$, since self-state transitions are the same in the direct and reversed chains.

7.1 Computation of p^r_{ij}

The transition probabilities in the time reversal chain (p^r_{ji}) satisfy:

$$\pi_i\, p_{ij} = \pi_j\, p^r_{ji} \tag{7.4}$$

Proof. Assume the chain in **steady state**. We have:

$$\begin{aligned} & P\{X(n+1) = j, X(n) = i\} = \\ & \quad P\{X^r(-n) = i, X^r(-n-1) = j\} = \\ & \quad P\{X^r(n+1) = i, X^r(n) = j\} \Rightarrow \\ & \quad P\{X(n) = i\}\, P\{X(n+1) = j \mid X(n) = i\} = \\ & \quad P\{X(n) = j\}\, P\{X^r(n+1) = i \mid X^r(n) = j\} \Rightarrow \\ & \quad \pi_i\, p_{ij} = \pi_j\, p^r_{ji}. \end{aligned}$$

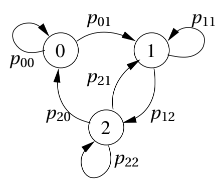
□

We can **compute** p^r_{ji} using the **reversed balance equations**:

$$\pi_i\, p_{ij} = \pi_j\, p^r_{ji} \Rightarrow \sum_{i\in U}\sum_{j\in V} \pi_i\, p_{ij} = \sum_{i\in U}\sum_{j\in V} \pi_j\, p^r_{ji} \Rightarrow$$

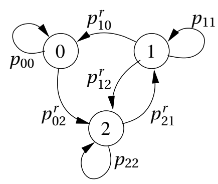
$$\boxed{F(U,V) = F^r(V,U)} \tag{7.5}$$

Example



$$\Rightarrow \begin{cases} \pi_0 = \dfrac{p_{12}\,p_{20}}{p_{12}\,p_{20} + p_{01}\,(p_{20} + p_{21}) + p_{01}\,p_{12}} \\ \pi_1 = \dfrac{p_{01}\,(p_{20} + p_{21})}{p_{12}\,p_{20} + p_{01}\,(p_{20} + p_{21}) + p_{01}\,p_{12}} \\ \pi_2 = \dfrac{p_{01}\,p_{12}}{p_{12}\,p_{20} + p_{01}\,(p_{20} + p_{21}) + p_{01}\,p_{12}} \end{cases}$$

Time reversal chain:



$$\Rightarrow \begin{cases} \pi_0\, p_{01} = \pi_1\, p^r_{10} \\ \pi_1\, p_{12} = \pi_2\, p^r_{21} \\ \pi_2\, p_{21} = \pi_1\, p^r_{12} \\ \pi_2\, p_{20} = \pi_0\, p^r_{02} \end{cases} \Rightarrow \begin{cases} p^r_{10} = \dfrac{p_{12}\, p_{20}}{p_{20}+p_{21}} \\ p^r_{21} = p_{20}+p_{21} \\ p^r_{12} = \dfrac{p_{12}\, p_{21}}{p_{20}+p_{21}} \\ p^r_{02} = p_{01} \end{cases}$$

Chapter 8
Reversible Chains

Definition

- A chain is reversible if:

$$\boxed{p_{ij} = p^r_{ij}} \tag{8.1}$$

- This equality implies the **reversibility balance equations**:

$$\boxed{\pi_i\, p_{ij} = \pi^r_i\, p^r_{ij} \Rightarrow F(U,V) = F^r(U,V)} \tag{8.2}$$

- Using both reversed ($F^r(U,V) = F(V,U)$) and reversibility balance equations, the following relation holds for a reversible chain (**detailed balance equations**):

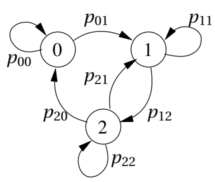
$$\boxed{F(U,V) = F(V,U)} \tag{8.3}$$

- NOTE: Compare with the **global balance equations**: $F(U,U^C) = F(U^C,U)$.

8.1 Kolmogorov Criteria

Definition of path

- Define a **path** as a possible sequence of transitions of the chain. For example, in the figure it could be $0 \rightsquigarrow 0 \rightsquigarrow 1 \rightsquigarrow 2$.



- We denote the **sequence of states** of one path l as:

$$(\mathbf{l},1) \rightsquigarrow (\mathbf{l},2) \rightsquigarrow \cdots \rightsquigarrow (\mathbf{l},m) \tag{8.4}$$

- For instance, if l is $0 \rightsquigarrow 0 \rightsquigarrow 1 \rightsquigarrow 2$, then $(\mathbf{l},1) = 0$, $(\mathbf{l},2) = 0$, $(\mathbf{l},3) = 1$, $(\mathbf{l},4) = 2$.
- We say that a path is **closed** if it starts and ends in the same state. For instance, a path stating and ending in state $(\mathbf{l},\mathbf{1})$:

$$(\mathbf{l},\mathbf{1}) \rightsquigarrow (l,2) \rightsquigarrow \cdots \rightsquigarrow (l,m) \rightsquigarrow (\mathbf{l},\mathbf{1}) \tag{8.5}$$

Kolmogorov Criteria

- Take a **closed path** l with $m \geq 0$ transitions, i.e.:

$$(l,1) \rightsquigarrow (l,2) \rightsquigarrow \cdots \rightsquigarrow (l,m) \rightsquigarrow (l,1),\ m \geq 0 \tag{8.6}$$

- The chain is **reversible iff for all** l:

$$\boxed{\begin{aligned} &p_{(l,1)(l,2)}\, p_{(l,2)(l,3)} \cdots p_{(l,m)(l,1)} = \\ &p_{(l,1)(l,m)}\, p_{(l,m)(l,m-1)} \cdots p_{(l,2)(l,1)} \end{aligned}} \tag{8.7}$$

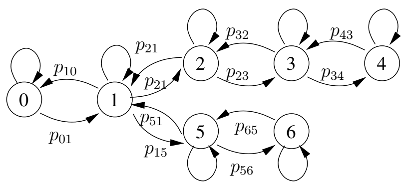
- **Proof**:
  - If the chain is reversible $\pi_i\, p_{ij} = \pi_j\, p_{ji}$ (detailed balance equations): $\Rightarrow \pi_{(l,k)}\, p_{(l,k)(l,k+1)} = \pi_{(l,k+1)}\, p_{(l,k+1)(l,k)}$
  - Multiplying for $k = 1,2\cdots m$ and simplifying we obtain the previous relation.

Corollary

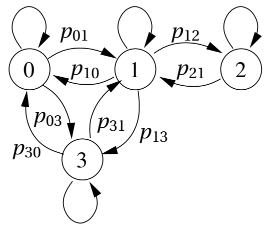
- A reversible chain must satisfy:

$$\boxed{\begin{aligned} p_{ij} > 0 \Rightarrow p_{ji} > 0 \\ p_{ij} = 0 \Rightarrow p_{ji} = 0 \end{aligned}} \tag{8.8}$$

- An **ergodic tree chain is always reversible**. We define a tree chain as chain having no cycles, i.e. the only possible transitions are bidirectional arcs between states, and self transitions.



Example



- The chain is **reversible iif**:

$$p_{01}\, p_{13}\, p_{30} = p_{10}\, p_{03}\, p_{31}$$

8.2 Product Form Solution

- Let $X(n)$ be a reversible MC with space state $S \Rightarrow$ the **stationary probabilities** of $X(n)$ can be computed as follows:
- Choose a state $\mathbf{s} \in S$,
- For every other state $\mathbf{i} \in S$, $i \neq s$ look for a possible path l_i from state s to state i:

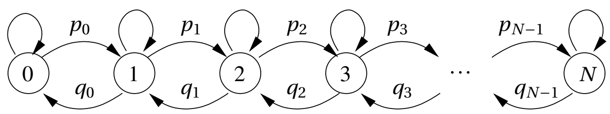
$$\mathbf{s} = (l_i,1) \rightsquigarrow (l_i,2) \rightsquigarrow \cdots \rightsquigarrow (l_i,m_{l_i}) = \mathbf{i},\ m_{l_i} \geq 1 \tag{8.9}$$

- The stationary probabilities are given by:

$$\boxed{\pi_i = \frac{\psi_i}{\sum_{j\in S} \psi_j},\ i \in S \quad \text{where } \psi_i = \begin{cases} 1, & i = s \\ \displaystyle\prod_{k=1}^{m_{l_i}-1} \frac{p_{(l_i,k)(l_i,k+1)}}{p_{(l_i,k+1)(l_i,k)}}, & i \neq s \end{cases}} \tag{8.10}$$

- **Proof** Use the detailed balance equations $\pi_i\, p_{ij} = \pi_j\, p_{ji}$.

8.3 Birth and Death Chains



- **Birth and death chains are reversible**.
- Applying the product form solution for reversible chains we obtain the general solution of birth death chains. Choosing $s = 0$:

$$\pi_i = \frac{\psi_i}{\sum_{j=0}^{N} \psi_j},\ i \geq 0 \quad \text{where } \psi_i = \begin{cases} 1, & i = 0 \\ \prod_{k=0}^{i-1} \frac{p_k}{q_k}, & i = 1, \cdots, N \end{cases} \tag{8.11}$$

Truncated Reversible Chain

- Consider a reversible MC X with a stationary distribution π_i.
- Suppose that **we truncate the chain** X and we obtain another irreducible chain X'.
- Then, X' is also reversible with stationary distribution:

$$\pi'_i = \frac{\pi_i}{G}, \quad \sum_k \pi'_k = 1 \tag{8.12}$$

Chapter 9
Research Example: Aloha

Access Protocol (see the paper of Kleinrock and Lam [5]).

- **Pure Aloha:**
  - Broadcast radio system.
  - **Single hop** system (all stations are in coverage).
  - Whenever a station has a frame ready, it is transmitted.
  - If two or more frames Tx overlap in time there is a **collision**, otherwise the frame is received correctly.
  - Colliding frames are reTx after a **random time** (**backoff**).
- **Slotted Aloha:**
  - Time is slotted.
  - Tx can only occur at the beginning of a slot.
  - Collisions occur when 2 or more stations Tx in the same slot.

9.1 Analysis with finite population

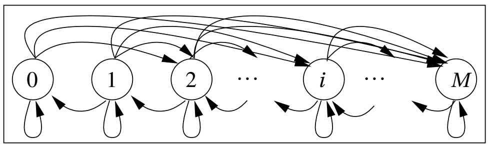
Assumptions

- **Slotted** Aloha.
- **Acks** are sent immediately after the reception of a frame, and are never lost.
- M nodes with a **buffer** of 1 frame.
- The **nodes** can be in 2 states:
  - Thinking: when the buffer is empty
  - Backlogged: when there is a frame in the buffer.
- A thinking node generate one frame in each slot with probability σ. When a frame collides, the frame is stored and the node becomes backlogged.
- A backlogged node ReTx the frame in each slot with probability ν.

Markov Chain

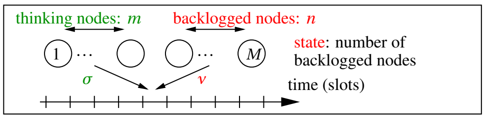
- The system **state**, $X(n)$, is the **number of backlogged nodes**:

$$p_{ij} = P\left(X(n) = j \text{ baklogged} \mid X(n-1) = i \text{ baklogged}\right) \tag{9.1}$$



Transition probabilities

- 0 for $j < i - 1$.
- for $j = i - 1$: no thinking Tx and only 1 backlogged Tx.
- for $j = i$:
  1. no thinking Tx and none or more than 1 backlogged Tx,
  2. only 1 thinking Tx and no backlogged Tx.
- for $j = i + 1$: 1 thinking and 1 or more backlogged Tx.
- for $j > i + 1$: $j - i$ thinking Tx, regardless of backlogged Tx.



In order to compute the previous events, define the probabilities:

- **Arrivals:** $Q_a(m,n)$, Probability of m thinking nodes Tx in a slot given that n nodes are backlogged:

$$Q_a(m,n) = P\left(\begin{smallmatrix} m \text{ think.} \\ \text{nodes Tx} \end{smallmatrix} \mid \begin{smallmatrix} n \text{nodes are} \\ \text{backlogged} \end{smallmatrix} \right) = \binom{M-n}{m} \sigma^m (1-\sigma)^{M-n-m} \tag{9.2}$$

- **Retransmissions:** $Q_r(m,n)$, Probability of m backlogged nodes Tx in a slot given that n nodes are backlogged:

$$Q_r(m,n) = P\left(\begin{smallmatrix} \mathbf{m} \text{ backl.} \\ \text{nodes Tx} \end{smallmatrix} \mid \begin{smallmatrix} \mathbf{n} \text{ nodes are} \\ \text{backlogged} \end{smallmatrix} \right) = \binom{n}{m} \nu^m (1-\nu)^{n-m} \tag{9.3}$$

- and we have:

$$p_{ij} = \begin{cases} 0, & j < i-1 \\ Q_a(0,i)\, Q_r(1,i), & j = i-1 \\ Q_a(0,i)\,(1 - Q_r(1,i)) + Q_a(1,i)\, Q_r(0,i), & j = i \\ Q_a(1,i)\,(1 - Q_r(0,i)), & j = i+1 \\ Q_a(j-i,i), & j > i+1 \end{cases} \tag{9.4}$$

9.1.1 Stationary distribution

- Solving the global balance equations:

$$\begin{aligned} \boldsymbol{\pi} &= \boldsymbol{\pi}\,\mathbf{P} \\ \boldsymbol{\pi}\,\mathbf{e} &= 1 \end{aligned} \tag{9.5}$$

- We obtain the probability of having i backlogged nodes:

$$\pi_i = P\left(i \text{ backlogged nodes}\right) \tag{9.6}$$

NOTE: there is **no closed form solution** of the chain. The matrix **P** must be constructed using the expression of p_{ij}, and solved numerically.

9.2 Throughput

- Define the probabilities:

$$P_{succ}(i) = P\left(\text{successful Tx} \mid i \text{ backlogged}\right) \tag{9.7}$$

- The **normalized throughput**, i.e. proportion of steps with a successful transmission) is:

$$S = \sum_{i=0}^{M} P\left(\text{successful Tx} \mid i \text{ backlogged}\right) \times P\left(i \text{ backlogged}\right) = \sum_{i=0}^{M} P_{succ}(i)\,\pi_i \tag{9.8}$$

- For a slot to be successful: (i) 1 thinking Tx and no backlogged Tx, or (ii) no thinking Tx and 1 backlogged Tx:

$$P_{succ}(i) = Q_a(1,i)\,Q_r(0,i) + Q_a(0,i)\,Q_r(1,i) \tag{9.9}$$

Notes on the throughput

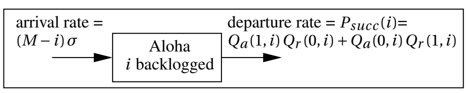
$$S = \sum_{i=0}^{M} P_{succ}(i)\,\pi_i \tag{9.10}$$

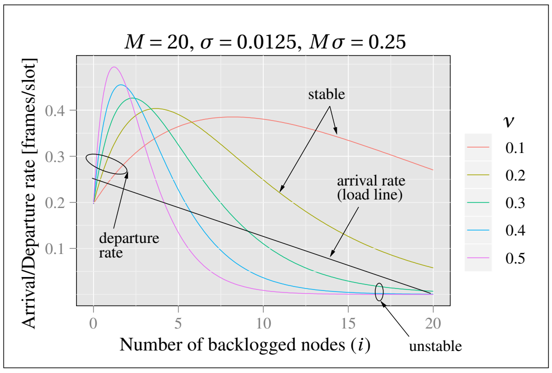
- For the **special case** $\sigma = \nu$ (thinking Tx with the same probability as backlogged): $P_{succ}(i) = M\sigma(1-\sigma)^{M-1}$, which does not depend on i, thus: $S = M\sigma(1-\sigma)^{M-1}$.
- The **offered load** (i.e. proportion of arrivals per slot) G is now: $G = M\sigma$, thus:

$$S = G\left(1 - \frac{G}{M}\right)^{M-1} \Rightarrow \lim_{M\to\infty} \mathbf{S} = \mathbf{G}\mathrm{e}^{-\mathbf{G}} \tag{9.11}$$

- We conclude that the **infinite population model** is the limit of the finite population if backlogged Tx with the same probability as thinking, and $M \to \infty$.

9.2.1 Dynamics





Note on the arrival rate (expected value of a binomial distribution):

$$\sum_{k=0}^{M-i} k \binom{M-i}{k} \sigma^k (1-\sigma)^{M-i-k} = (\mathbf{M}-\mathbf{i})\,\sigma$$

Solving the chain: $S = \sum_{i=0}^{M} P_{succ}(i)\,\pi_i$

ν	S
0.1	2.38e-01
0.2	2.42e-01
0.3	**1.30e-02**
0.4	**4.98e-04**
0.5	**1.90e-05**

9.2.2 Stabilizing Aloha

- The **retransmission probabilities** must adapt in accordance with the state of the system.
- Example: **binary exponential backoff** (ethernet). The retransmission rate at retransmission i is adapted as $\nu = 2^{-i}$. Thus, the higher are the number of retransmission trials i, the lower (exponentially) is the retransmission rate.

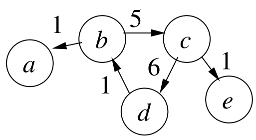
Chapter 10
Finite Absorbing Chains

Canonical Form

- Let $\mathbf{P}^{rxr}$ be the transition probability matrix of a chain with r states: s **transient** states and $r-s$ **absorbing** states. We can write $\mathbf{P}^{rxr}$ in the **canonical** form:

$$\mathbf{P}^{rxr} = \begin{bmatrix} \mathbf{Q}^{s\times s} & \mathbf{R}^{s\times r-s} \\ \mathbf{0}^{r-s\times s} & \mathbf{I}^{r-s\times r-s} \end{bmatrix} \tag{10.1}$$

Example



$$\mathbf{P} = \begin{array}{c|c} & \begin{array}{ccc|cc} b & c & d & a & e \end{array} \\ \begin{array}{c} b \\ c \\ d \\ a \\ e \end{array} & \left[\begin{array}{ccc|cc} 0 & 0.1 & 0 & 0.9 & 0 \\ 0 & 0 & 0.8 & 0 & 0.2 \\ 0.3 & 0 & 0.7 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right] \end{array}$$

10.1 Results

- Define:

$$\mathbf{n_{ij}} = \left\{\begin{matrix}\text{number of steps in state } j \text{ before} \\ \text{absorption, starting from state } i\end{matrix}\right\},$$
$$\mathbf{t_i} = \left\{\begin{matrix}\text{number of steps in transient states before} \\ \text{absorption, starting from state } i\end{matrix}\right\}, \tag{10.2}$$
$$\mathbf{b_{ij}} = P\big(\text{probability to be absorbed } j \text{ starting } i\big)$$

- Then:

$$\begin{aligned}\{\mathrm{E}[n_{ij}]\} &= \mathbf{N} = (\mathbf{I}-\mathbf{Q})^{-1}\\ \{\mathrm{Var}[n_{ij}]\} &= \mathbf{N}\,(2\mathbf{N}_{\text{diag}} - \mathbf{I}) - \mathbf{N}\text{sqr}\\ \{\mathrm{E}[t_i]\} &= \boldsymbol{\tau} = \mathbf{N}\,\mathbf{e}\\ \{\mathrm{Var}[t_i]\} &= (2\mathbf{N}-\mathbf{I})\,\boldsymbol{\tau} - \boldsymbol{\tau}_{\text{sqr}}\\ \{b_{ij}\} &= \mathbf{B} = \mathbf{N}\mathbf{R}.\end{aligned} \tag{10.3}$$

where $\{a_{ij}\}$ is a matrix with a_{ij} as element ij and $\mathbf{e}$ is a column vector of 1s. $\mathbf{N}$ is called the **fundamental matrix**.

Proofs

- $\{\mathrm{E}[n_{ij}]\} = \mathbf{N} = (\mathbf{I}-\mathbf{Q})^{-1}$

Proof.
$$\begin{aligned}\mathrm{E}[n_{ij}] &= \sum_{k\in A} p_{ik}\delta_{ij} + \sum_{k\in T} p_{ik}\mathrm{E}[n_{kj}+\delta_{ij}] =\\ &\qquad \delta_{ij} + \sum_{k\in T} p_{ik}\mathrm{E}[n_{kj}]\\ &\qquad \Rightarrow \{\mathrm{E}[n_{ij}]\} = \mathbf{N} = \mathbf{I} + \mathbf{Q}\,\mathbf{N} \Rightarrow \mathbf{N} = (\mathbf{I}-\mathbf{Q})^{-1}\end{aligned}$$

where A is the set of absorbing states and T is the set of transient states.

Notation: $\delta_{ij} = I(i=j) = \begin{cases}1, & i=j,\\ 0, & \text{otherwise.}\end{cases}$ □

- $\{\mathrm{Var}[n_{ij}]\} = \mathbf{N}\,(2\mathbf{N}_{\text{diag}} - \mathbf{I}) - \mathbf{N}_{\text{sqr}}$

Proof.
$$\mathrm{Var}[n_{ij}] = \mathrm{E}[n_{ij}^2] - \mathrm{E}[n_{ij}]^2 \Rightarrow$$
$$\{\mathrm{Var}[\mathbf{n_{ij}}]\} = \{\mathrm{E}[\mathbf{n_{ij}^2}]\} - \mathbf{N}_{\text{sqr}}$$
$$\begin{aligned}\mathrm{E}[\mathbf{n_{ij}^2}] &= \sum_{k\in A} p_{ik}\delta_{ij}^2 + \sum_{k\in T} p_{ik}\mathrm{E}[(n_{kj}+\delta_{ij})^2] =\\ &\sum_{k\in A} p_{ik}\delta_{ij} + \sum_{k\in T} p_{ik}(\mathrm{E}[n_{kj}^2] + 2\mathrm{E}[n_{kj}]\delta_{ij} + \delta_{ij}) =\\ &\qquad \delta_{ij} + \sum_{k\in T}(p_{ik}\mathrm{E}[n_{kj}^2] + 2\,p_{ik}\mathrm{E}[n_{kj}]\delta_{ij}) \Rightarrow\end{aligned}$$
$$\begin{aligned}\{\mathrm{E}[\mathbf{n_{ij}^2}]\} &= \mathbf{I} + \mathbf{Q}\,\{\mathrm{E}[n_{ij}^2]\} + 2\,(\mathbf{Q}\mathbf{N})_{\text{diag}} =\\ &(\mathbf{I}-\mathbf{Q})^{-1}(\mathbf{I} + 2(\mathbf{Q}\mathbf{N})_{\text{diag}}) =\\ &\mathbf{N}(\mathbf{I} + 2(\mathbf{N}-\mathbf{I})_{\text{diag}}) = \mathbf{N}(2\mathbf{N}_{\text{diag}} - \mathbf{I})\end{aligned}$$ □

- $\{\mathrm{E}[t_i]\} = \boldsymbol{\tau} = \mathbf{N}\mathbf{e}$

Proof.
$$\mathrm{E}[t_i] = \sum_{k\in T}\mathrm{E}[n_{ij}]\} \Rightarrow \{E[t_i]\} = \boldsymbol{\tau} = \mathbf{N}\mathbf{e}$$ □

- $\{\mathrm{Var}[t_i]\} = (2\mathbf{N}-\mathbf{I})\,\boldsymbol{\tau} - \boldsymbol{\tau}_{\text{sqr}}$

Proof.
$$\mathrm{Var}[t_i] = \mathrm{E}[t_i^2] - \mathrm{E}[t_i]^2 \Rightarrow \{\mathrm{Var}[\mathbf{t_i}]\} = \{\mathrm{E}[\mathbf{t_i^2}]\} - \tau_{\text{sqr}}$$
$$\begin{aligned}\mathrm{E}[\mathbf{t_i^2}] &= \sum_{k\in A} p_{ik} + \sum_{k\in T} p_{ik}\mathrm{E}[(t_k+1)^2] =\\ &\sum_{k\in A} p_{ik} + \sum_{k\in T} p_{ik}(\mathrm{E}[t_k^2] + 2\mathrm{E}[t_k] + 1) =\\ &\qquad 1 + \sum_{k\in T}(p_{ik}\mathrm{E}[t_k^2] + 2\,p_{ik}\mathrm{E}[t_k]) \Rightarrow\end{aligned}$$
$$\begin{aligned}\{\mathrm{E}[\mathbf{t_i^2}]\} &= \mathbf{e} + \mathbf{Q}\,\{\mathrm{E}[t_i^2]\} + 2\,\mathbf{Q}\,\tau =\\ &(\mathbf{I}-\mathbf{Q})^{-1}(\mathbf{e} + 2\,\mathbf{Q}\,\tau) =\\ &\mathbf{N}\,(\mathbf{e} + 2\,\mathbf{Q}\,\tau) = \tau + 2\mathbf{N}\mathbf{Q}\,\tau =\\ &\tau + 2\,(\mathbf{N}-\mathbf{I})\tau = (2\mathbf{N}-\mathbf{I})\tau\end{aligned}$$ □

- $\{b_{ij}\} = \mathbf{B} = \mathbf{N}\mathbf{R}$

Proof.
$$b_{ij} = p_{ij} + \sum_{k\in T} p_{ik}b_{kj},\ j \in A \Rightarrow$$
$$\{b_{ij}\} = \mathbf{B} = \mathbf{R} + \mathbf{Q}\mathbf{B}$$
$$\Rightarrow \mathbf{B} = (\mathbf{I}-\mathbf{Q})^{-1}\mathbf{R} = \mathbf{N}\mathbf{R}.$$ □

10.2 Extension of the Results

- The previous results can be generalized to any group of states of $\mathbf{P}$:
- A set S is referred to as **open** if the chain can reach some state of S^c starting from any state of S. Let

$$\mathbf{Q} = \{p_{ij}, i \in S, j \in S\} \tag{10.4}$$
$$\mathbf{R} = \{p_{ij}, i \in S, j \in S^c\} \tag{10.5}$$

Let assume that the process starts from $i \in S$. Define:

$$\mathbf{n_{ij}} = \left\{\begin{matrix}\text{number of steps in state } j \text{ before} \\ \text{leaving } S\text{, starting from state } i\end{matrix}\right\},$$
$$\Rightarrow \{\mathrm{E}[n_{ij}]\} = \mathbf{N} = (\mathbf{I}-\mathbf{Q})^{-1}.$$

- Similarly for the other results, e.g. $\boldsymbol{\tau} = \{\mathrm{E}[t_i]\} = \mathbf{N}\mathbf{e}$ and $\mathbf{B} = \{b_{ij}\} = \mathbf{N}\mathbf{R}$.

10.3 Inverse of a matrix

Cofactors

$$\mathbf{A}^{-1} = \frac{1}{\det\mathbf{A}}\mathbf{C}^{\mathrm{T}} \tag{10.6}$$

where $\mathbf{C}^{\mathrm{T}}$ is the transposed cofactor matrix: $c_{ij} = (-1)^{i+j}\det\mathbf{M}_{ij}$, and $\mathbf{M}_{ij}$ are the minor matrices obtained removing the row i and column j from $\mathbf{A}$.

Gaussian Elimination

Do the transformation:

$$[\mathbf{A}\,|\,\mathbf{I}] \to \left[\mathbf{I}\,|\,\mathbf{A}^{-1}\right] \tag{10.7}$$

using the elementary row operations:

- Swapping two rows.
- Multiplying a row by a nonzero number.
- Adding a multiple of one row to another row.

Bibliography

[1] Llorenç Cerdà-Alabern. *Transient Solution of Markov Chains Using the Uniformized Vandermonde Method*. Tech. rep. UPC-DAC-RR-XCSD-2010-2. Universitat Politècnica de Catalunya, Dec. 2010. URL: https://www.ac.upc.edu/app/research-reports/html/research_center_index-XCSD-2010,en.html.

[2] William Feller. *An Introduction to Probability Theory and Its Applications: Volume One*. 3rd ed. John Wiley & Sons, 1968.

[3] Mor Harchol-Balter. *Performance Modeling and Design of Computer Systems: Queueing Theory in Action*. Cambridge University Press, 2013.

[4] John G Kemeny and James Laurie Snell. *Finite markov chains*. Vol. 356. van Nostrand Princeton, NJ, 1960.

[5] Leonard Kleinrock and Simon Lam. "**Packet Switching in a Multiaccess Broadcast Channel: Performance Evaluation**". In: *Communications, IEEE Transactions on* 23.4 (1975), pp. 410–423.

[6] Randolph Nelson. *Probability, stochastic processes, and queueing theory: the mathematics of computer performance modeling*. Springer, 1995.

[7] Kishor S Trivedi. *Probability & statistics with reliability, queuing and computer science applications*. John Wiley & Sons, 2008.

Table with some distributions

Distribution	Parametres	Density	Mean	Variance	Characteristic Function
Bernoulli	$0 \le p \le 1$ $q = 1-p$	$p^k(1-p)^{1-k}$ $k = 0,1$	p	$p(1-p)$	$q + p\,\mathrm{e}^{i\,t}$
Binomial	$0 \le p \le 1$ $q = 1-p$	$\binom{n}{k} p^k(1-p)^{n-k}$ $k = 0,1,\dots n$	$n\,p$	$n\,p(1-p)$	$(q + p\,\mathrm{e}^{i\,t})^n$
Geometric	$0 \le p \le 1$ $q = 1-p$	$p(1-p)^k$ $k \ge 0$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$	$\frac{p}{1-q\,\mathrm{e}^{i\,t}}$
Negative binomial	$r > 0$ $0 \le p \le 1$ $q = 1-p$	$\binom{k+r-1}{k} p^r q^k$ $k \ge 0$	$r\,\frac{1-p}{p}$	$r\,\frac{1-p}{p^2}$	$\left(\frac{p}{1-q\,\mathrm{e}^{i\,t}}\right)^r$
Poisson	$\lambda > 0$	$\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}, \quad k \ge 0$	λ	λ	$\exp\left\{\lambda(\mathrm{e}^{i\,t}-1)\right\}$
Normal $N(\mu,\sigma)$	$\mu \in \mathbb{R}$ $\sigma > 0$	$\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}$ $x \in \mathbb{R}$	μ	σ^2	$\exp\left\{\mu\,i\,t - \frac{t^2\sigma^2}{2}\right\}$
Uniform	$a < b$	$\frac{1}{b-a}, \quad a \ge x \ge b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{\mathrm{e}^{i\,t\,b}-\mathrm{e}^{i\,t\,a}}{i\,t(b-a)}$
Exponential	α	$\alpha\,\mathrm{e}^{-\alpha\,x}, \quad x \ge 0$	$\frac{1}{\alpha}$	$\frac{1}{\alpha^2}$	$\left(1-\frac{i\,t}{\alpha}\right)^{-1}$
Gamma $\gamma(n,\alpha)$	$\alpha > 0$, $n > 0$	$\frac{\alpha^n x^{n-1}\mathrm{e}^{-\alpha\,x}}{\Gamma(n)}, \quad x \ge 0$	$\frac{n}{\alpha}$	$\frac{n}{\alpha^2}$	$\left(1-\frac{i\,t}{\alpha}\right)^{-n}$
Beta $\beta(p,q)$	$p > 0$, $q > 0$	$\frac{x^{p-1}(1-x)^{q-1}}{\mathrm{B}(p,q)}$, $0 \ge x \ge 1$	$\frac{p}{p+q}$	$\frac{p\,q}{(p+q)^2(p+q+1)}$	

$$\Gamma(x) = \int_0^\infty e^{-t}\,t^{x-1}\,dt, \Gamma(n) = (n-1)! \quad \mathrm{B}(p,q) = \int_0^1 t^{p-1}(1-t)^{q-1}\,dt, \mathrm{B}(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$$