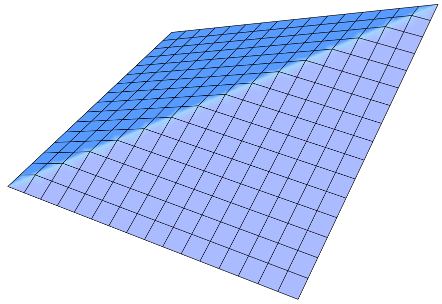


STAT 433: Stochastic Processes

Prof. Yi Shen • Fall 2014 • University of Waterloo
MW 2:30–3:50PM in MC 4061

Transcribed, LaTeXed, and edited by Rob Zimmerman

Note: These are personal lecture notes, not official course materials. They may contain errors (which by default you should attribute to me, Rob Zimmerman, rather than the course instructor). The permanent home of these — and all of my other lecture notes — is https://rob-zimmerman.github.io/coursenotes.

Contents

Lecture 1 — Monday, September 08, 2014

1. Discrete-Time Markov Chains

Definitions and Transition Matrices

We start off by reviewing some familiar concepts. A **random variable** is a real-valued function on a sample space, $X : \Omega \to \mathbb{R}$. A **stochastic process** is a family of random variables $\{X_t\}_{t\in T}$ defined on a common sample space Ω. The index set T often represents time, although this need not always be the interpretation. In discrete time, the usual index set is $T = \{0, 1, 2, \dots\}$ and we write X_n for $n = 0, 1, 2, \dots$, so that

$$\{X_n : n = 0, 1, 2, \dots\} = \{X_n\}_{n=0,1,2,\dots}.$$

In continuous time, we often use $T = \{t \mid t \geq 0\}$ and write

$$\{X_t : t \geq 0\} = \{X_t\}_{t \geq 0}.$$

The **state space** S is the collection of all possible values of the random variables in the process.

Definition 1.1 A stochastic process $\{X_n : n = 0, 1, 2, \dots\}$ is called a **discrete-time Markov chain** if the following hold:

1. The state space S is at most countable, meaning that it is either finite or countably infinite.
2. It satisfies the **Markov property**: for any $n = 0, 1, 2, \dots$,

$$\mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n, X_{n-1} = x_{n-1}, \dots) = \mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n).$$

whenever the conditioning events have positive probability. In other words, the future depends only on the present state.

Definition 1.2 The **transition probability** from state $i \in S$ at time n to state $j \in S$ at time $n + 1$ is given by

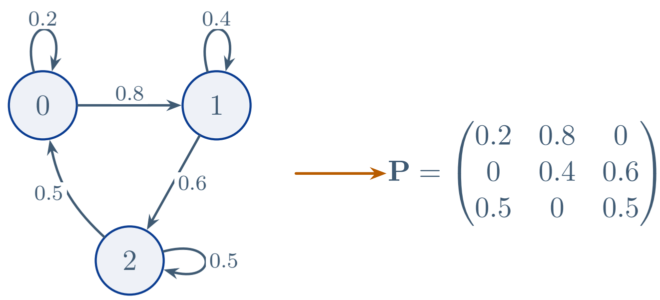
$$\mathbb{P}_{n,i,j} = \mathbb{P}(X_{n+1} = j \mid X_n = i)$$

for $n = 0, 1, 2, \dots$. In this course, we focus on the case where $\mathbb{P}_{n,i,j}$ does not depend on n, and simply write $\mathbb{P}_{i,j}$. In this case, the Markov chain is called **time-homogeneous**, or simply **homogeneous**. This will be the default discrete-time Markov chain setting. The matrix $\mathbf{P} = \{\mathbb{P}_{i,j}\}_{i,j \in S}$ is called the **one-step transition probability matrix**.

The same one-step information can be read from a state diagram or from a transition matrix. In Figure 1, each directed edge from i to j corresponds to the entry in row i and column j.

The transition matrix satisfies

$$\mathbb{P}_{i,j} \geq 0 \quad \text{for all } i, j \in S$$



outgoing edge weights from each state form one row of **P**

Figure 1

and

$$\sum_{j \in S} \mathbb{P}_{i,j} = 1,$$

so the entries in each row add to 1. The n-step transition probability is

$$\mathbb{P}_{i,j}^{(n)} = \mathbb{P}(X_{n+m} = j \mid X_m = i) = \mathbb{P}(X_n = j \mid X_0 = i)$$

and the corresponding matrix is $\mathbf{P}^{(n)} = \{\mathbb{P}_{i,j}^{(n)}\}_{i,j \in S}$. The one-step and n-step transition matrices are related by the Chapman–Kolmogorov equation.

Theorem 1.3 (Chapman–Kolmogorov equation) $\mathbf{P}^{(n)} = \mathbf{P}^n$.

Proof. We prove the result by induction. The case $n = 1$ follows from the definition, and $\mathbf{P}^{(0)} = \mathbf{I} = \mathbf{P}^0$. Assume that $\mathbf{P}^{(n)} = \mathbf{P}^n$. Then

$$\begin{aligned}
\mathbb{P}_{i,j}^{(n+1)} &= \mathbb{P}(X_{n+1} = j \mid X_0 = i) \\
&= \sum_{k \in S} \mathbb{P}(X_{n+1} = j, X_n = k \mid X_0 = i) \\
&= \sum_{k \in S} \mathbb{P}(X_{n+1} = j \mid X_n = k, X_0 = i)\mathbb{P}(X_n = k \mid X_0 = i) \\
&= \sum_{k \in S} \mathbb{P}(X_{n+1} = j \mid X_n = k)\mathbb{P}(X_n = k \mid X_0 = i) \\
&= \sum_{k \in S} \mathbb{P}_{k,j}\mathbb{P}_{i,k}^{(n)} \\
&= (\mathbf{P}^{(n)}\mathbf{P})_{i,j} \\
&= (\mathbf{P}^n\mathbf{P})_{i,j} = (\mathbf{P}^{n+1})_{i,j}.
\end{aligned}$$

□

In particular, this implies that $\mathbf{P}^{(n)} = \mathbf{P}^{(m)} \cdot \mathbf{P}^{(n-m)}$ for $0 \le m \le n$, or equivalently

$$\mathbb{P}_{i,j}^{(n)} = \sum_{k \in S} \mathbb{P}_{i,k}^{(m)} \mathbb{P}_{k,j}^{(n-m)} \quad \text{for all } i, j \in S,\ 0 \le m \le n.$$

This is the Chapman–Kolmogorov equation.

Let $\boldsymbol{\alpha}_n = (\alpha_{n,0}, \alpha_{n,1}, \dots)^\top$ be the distribution of X_n, where $\alpha_{n,k} = \mathbb{P}(X_n = k)$ for all k. Then $\alpha_{n,k} \ge 0$ for all n, k, and

$$\sum_{k \in S} \alpha_{n,k} = 1$$

for all $n = 0, 1, 2, \dots$. The distribution

$$\boldsymbol{\alpha}_0 = (\mathbb{P}(X_0 = 0), \mathbb{P}(X_0 = 1), \mathbb{P}(X_0 = 2), \dots)^\top$$

is called the initial distribution.

Theorem 1.4 $\boldsymbol{\alpha}_n = (\mathbf{P}^\top)^n \boldsymbol{\alpha}_0$.

Proof. For $j \in S$,

$$\alpha_{n,j} = \mathbb{P}(X_n = j) = \sum_{i \in S} \mathbb{P}(X_n = j \mid X_0 = i)\mathbb{P}(X_0 = i) = \sum_{i \in S} ((\mathbf{P}^\top)^n)_{j,i}\, \alpha_{0,i} = \Big((\mathbf{P}^\top)^n \boldsymbol{\alpha}_0\Big)_j .$$

□

Lecture 2 — Wednesday, September 10, 2014

The probabilistic behaviour of a homogeneous discrete-time Markov chain is determined by its initial distribution and its transition probability matrix. We saw that $\mathbf{P}^{(n)} = \mathbf{P}^n$, and so $\mathbf{P}^{(n)} = \mathbf{P}^{(m)}\mathbf{P}^{(n-m)}$.

More generally, for any $n \in \mathbb{N}$, the finite-dimensional distribution is

$$\begin{aligned}\mathbb{P}(X_n = x_n, \dots, X_0 = x_0) &= \mathbb{P}(X_0 = x_0)\mathbb{P}(X_1 = x_1 \mid X_0 = x_0)\mathbb{P}(X_2 = x_2 \mid X_1 = x_1) \\ &\quad \cdots \mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1}) \\ &= \alpha_{0,x_0}\mathbb{P}_{x_0,x_1}\mathbb{P}_{x_1,x_2}\cdots\mathbb{P}_{x_{n-1},x_n}\end{aligned}$$

by repeatedly applying the Markov property. Even more generally, if

$$0 \le t_1 < \cdots < t_n,$$

then

$$\mathbb{P}(X_{t_n} = x_{t_n}, \dots, X_{t_1} = x_{t_1}) = \mathbb{P}(X_{t_1} = x_{t_1})(\mathbf{P}^{t_2-t_1})_{x_{t_1},x_{t_2}}\cdots(\mathbf{P}^{t_n-t_{n-1}})_{x_{t_{n-1}},x_{t_n}},$$

where

$$\mathbb{P}(X_{t_1} = x_{t_1}) = \alpha_{t_1,x_{t_1}} = \left((\mathbf{P}^\top)^{t_1}\boldsymbol{\alpha}_0\right)_{x_{t_1}} = \sum_{k\in S}\alpha_{0,k}(\mathbf{P}^{t_1})_{k,x_{t_1}}.$$

This again shows that the probabilistic properties of a homogeneous discrete-time Markov chain are fully characterized by the initial distribution and the transition matrix $\mathbf{P}$.

Classification of States

Definition 2.1 State j is **accessible** from state i if there exists $n \in \{0, 1, 2, \dots\}$ such that $\mathbb{P}^{(n)}_{i,j} > 0$. Otherwise, j is not accessible from i. If, in addition, i is accessible from j, then i and j **communicate**, and in this case we write $i \leftrightarrow j$.

Intuitively, if state j is accessible from state i, then it is possible to reach j from i in a finite number of steps.

The communication relation $\leftrightarrow$ is an equivalence relation. It is reflexive since $\mathbb{P}^{(0)}_{i,i} = 1 > 0$, and it is symmetric by definition. It is also transitive: if $i \leftrightarrow j$ and $j \leftrightarrow k$, then there exist m, n such that

$$\mathbb{P}^{(m+n)}_{i,k} = \sum_{l\in S}\mathbb{P}^{(m)}_{i,l}\mathbb{P}^{(n)}_{l,k} \ge \mathbb{P}^{(m)}_{i,j}\mathbb{P}^{(n)}_{j,k} > 0.$$

Thus k is accessible from i, and the reverse accessibility follows similarly. Therefore $i \leftrightarrow k$. The relation $\leftrightarrow$ partitions the state space into disjoint equivalence classes: states that communicate with each other all belong to one class. These classes form a partition of S, meaning that $S_i \cap S_j = \emptyset$ for $i \neq j$ and $\bigcup_i S_i = S$.

We can find the equivalence classes by drawing a directed graph whose nodes are the states and whose directed edges are $i \to j$ whenever $\mathbb{P}_{i,j} > 0$.

Definition 2.2 A Markov chain with only one equivalence class is called **irreducible**.

Definition 2.3 The number

$$d_{(i)} := \gcd\{n \geq 1 \mid \mathbb{P}_{i,i}^{(n)} > 0\}$$

is called the **period** of state i. If $\mathbb{P}_{i,i}^{(n)} = 0$ for all $n > 0$, we set $d_{(i)} = \infty$. The state is called **aperiodic** if $d_{(i)} = 1$. The Markov chain is called **aperiodic** if all of its states are aperiodic.

Theorem 2.4 If $i \leftrightarrow j$, then $d_{(i)} = d_{(j)}$.

Proof. Since $i \leftrightarrow j$, there exist m, n such that $\mathbb{P}_{i,j}^{(m)} > 0$ and $\mathbb{P}_{j,i}^{(n)} > 0$. For any l such that $\mathbb{P}_{j,j}^{(l)} > 0$,

$$\mathbb{P}_{i,i}^{(m+l+n)} \geq \mathbb{P}_{i,j}^{(m)}\mathbb{P}_{j,j}^{(l)}\mathbb{P}_{j,i}^{(n)} > 0.$$

On the other hand,

$$\mathbb{P}_{i,i}^{(m+n)} \geq \mathbb{P}_{i,j}^{(m)}\mathbb{P}_{j,i}^{(n)} > 0.$$

Since $d_{(i)}$ divides both $m+n$ and $m+n+l$, it divides l. Therefore $d_{(i)}$ divides $\gcd\{l \mid \mathbb{P}_{j,j}^{(l)} > 0\} = d_{(j)}$. By symmetry, $d_{(j)}$ divides $d_{(i)}$, and hence $d_{(i)} = d_{(j)}$. $\square$

Consider the Markov chain represented by the directed graph in Figure 2:

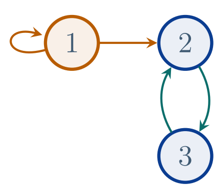


FIGURE 2

State 1 has a self-loop, so $d_{(1)} = 1$, while states 2 and 3 communicate with each other and have period 2. More generally, a state may have period 1 even when its one-step return probability is 0: it is enough, for example, to have possible return times whose greatest common divisor is 1. We thus make the following remarks:

Remark 2.5

1. In general, $d_{(i)} = k \not\Longrightarrow \mathbb{P}_{i,i}^{(k)} > 0$.
2. If the Markov chain is irreducible, all states have the same period. In this case, we also call it the period of the Markov chain.

Recurrence and Transience

For $n \in \mathbb{Z}^+$, define

$$f_{i,j}^{(n)} := \mathbb{P}(X_n = j, X_{n-1} \neq j, \ldots, X_1 \neq j \mid X_0 = i).$$

Intuitively, this is the probability that the chain visits state j for the first time at time n, given that it starts from state i.

The quantities $\mathbb{P}_{i,j}^{(n)}$ and $f_{i,j}^{(n)}$ are related as follows. First, $\mathbb{P}_{i,j}^{(n)} \geq f_{i,j}^{(n)}$. Also,

$$\sum_{n=1}^{\infty} f_{i,j}^{(n)} = \mathbb{P}(\text{the Markov chain ever visits } j \mid X_0 = i) \leq 1.$$

More precisely,

$$\mathbb{P}_{i,j}^{(n)} = \sum_{k=1}^{n} f_{i,j}^{(k)} \mathbb{P}_{j,j}^{(n-k)} = f_{i,j}^{(n)} + \sum_{k=1}^{n-1} f_{i,j}^{(k)} \mathbb{P}_{j,j}^{(n-k)}.$$

Lecture 3 — Monday, September 15, 2014

Definition 3.1 Let

$$\begin{aligned} f_{i,j}^{(n)} &:= \mathbb{P}(X_n = j, X_{n-1} \neq j, \ldots, X_1 \neq j \mid X_0 = i) \\ &= \mathbb{P}(\text{the Markov chain visits } j \text{ first on the } n\text{-th step} \mid X_0 = i). \end{aligned}$$

We know that $\mathbb{P}_{i,j}^{(n)} \geq f_{i,j}^{(n)}$ and

$$\sum_{n=1}^{\infty} f_{i,j}^{(n)} \leq 1.$$

The relation between $\mathbb{P}_{i,j}^{(n)}$ and $f_{i,j}^{(n)}$ is this:

Theorem 3.2

$$\mathbb{P}_{i,j}^{(n)} = \sum_{k=1}^{n} f_{i,j}^{(k)} \mathbb{P}_{j,j}^{(n-k)} = f_{i,j}^{(n)} + \sum_{k=1}^{n-1} f_{i,j}^{(k)} \mathbb{P}_{j,j}^{(n-k)}.$$

Proof.

$$\begin{aligned}
\mathbb{P}_{i,j}^{(n)} &= \mathbb{P}(X_n = j \mid X_0 = i) \\
&= \sum_{k=1}^{n} \mathbb{P}(X_n = j,\ X \text{ first visits } j \text{ at time } k \mid X_0 = i) \\
&= \sum_{k=1}^{n} \mathbb{P}(X_n = j \mid X \text{ first visits } j \text{ at time } k,\ X_0 = i)\mathbb{P}(X \text{ first visits } j \text{ at time } k \mid X_0 = i) \\
&= \sum_{k=1}^{n} \mathbb{P}(X_n = j \mid X_k = j) f_{i,j}^{(k)} \\
&= \sum_{k=1}^{n} f_{i,j}^{(k)} \mathbb{P}_{j,j}^{(n-k)} \\
&= f_{i,j}^{(n)} + \sum_{k=1}^{n-1} f_{i,j}^{(k)} \mathbb{P}_{j,j}^{(n-k)}.
\end{aligned}$$

$\square$

We can rewrite this as

$$f_{i,j}^{(n)} = \mathbb{P}_{i,j}^{(n)} - \sum_{k=1}^{n-1} f_{i,j}^{(k)} \mathbb{P}_{j,j}^{(n-k)},$$

which is a recursive formula to calculate $f_{i,j}^{(n)}$.

Definition 3.3 We define

$$f_{i,j} := \mathbb{P}(\text{the Markov chain ever goes to state } j \mid X_0 = i) = \sum_{n=1}^{\infty} f_{i,j}^{(n)}.$$

Definition 3.4 State i is called **transient** if $f_{i,i} < 1$. Otherwise, state i is called **recurrent** (i.e., when $f_{i,i} = 1$).

Another way to characterize recurrence and transience is as follows. Define

$$M_i = \text{total number of times the Markov chain visits state } i \text{ after time } 0 = \sum_{n=1}^{\infty} \mathbb{1}_{\{X_n = i\}}.$$

If $f_{i,i} < 1$, then

$$\mathbb{P}(M_i = k \mid X_0 = i) = f_{i,i}^k (1 - f_{i,i}), \quad k = 0, 1, 2, \ldots.$$

Thus, given $X_0 = i$, M_i follows a geometric distribution with parameter $1 - f_{i,i}$, and

$$\mathbb{E}\left[M_i \mid X_0 = i\right] = \frac{f_{i,i}}{1 - f_{i,i}} < \infty$$

when $f_{i,i} < 1$. On the other hand, notice that

$$\lim_{f_{i,i} \to 1} \frac{f_{i,i}}{1 - f_{i,i}} = \infty.$$

Alternatively,

$$\mathbb{E}\left[M_i \mid X_0 = i\right] = \sum_{n=1}^{\infty} \mathbb{P}(X_n = i \mid X_0 = i) = \sum_{n=1}^{\infty} \mathbb{P}_{i,i}^{(n)}.$$

If $f_{i,i} = 1$, then $\mathbb{E}[M_i \mid X_0 = i] = \infty$. Therefore i is recurrent if and only if $\mathbb{E}[M_i \mid X_0 = i] = \infty$, and i is transient if and only if $\mathbb{E}[M_i \mid X_0 = i] < \infty$.

Remark 3.5 Conditional on $X_0 = i$, we actually have recurrence if and only if $M_i = \infty$ almost surely, and transience if and only if $M_i < \infty$ almost surely.

Yet another way to see the situation is to observe that

$$\begin{aligned}
\mathbb{E}\left[M_i \mid X_0 = i\right] &= \mathbb{E}\left[\sum_{n=1}^{\infty} \mathbb{1}_{\{X_n = i\}} \mid X_0 = i\right] \\
&= \sum_{n=1}^{\infty} \mathbb{E}\left[\mathbb{1}_{\{X_n = i\}} \mid X_0 = i\right] \\
&= \sum_{n=1}^{\infty} \mathbb{P}(X_n = i \mid X_0 = i)
\end{aligned}$$

$$= \sum_{n=1}^{\infty} \mathbb{P}_{i,i}^{(n)}.$$

Thus i is recurrent if and only if

$$\sum_{n=1}^{\infty} \mathbb{P}_{i,i}^{(n)} = \infty.$$

Recurrence and transience are class properties.

Theorem 3.6 If $i \leftrightarrow j$ and i is recurrent, then j is also recurrent.

Proof. Since $i \leftrightarrow j$, there exist m, n such that $\mathbb{P}_{i,j}^{(m)} > 0$ and $\mathbb{P}_{j,i}^{(n)} > 0$. Then

$$\sum_{r=1}^{\infty} \mathbb{P}_{j,j}^{(r)} \geq \sum_{l=1}^{\infty} \mathbb{P}_{j,j}^{(n+l+m)} \geq \sum_{l=1}^{\infty} \mathbb{P}_{j,i}^{(n)} \mathbb{P}_{i,i}^{(l)} \mathbb{P}_{i,j}^{(m)} = \mathbb{P}_{j,i}^{(n)} \mathbb{P}_{i,j}^{(m)} \sum_{l=1}^{\infty} \mathbb{P}_{i,i}^{(l)}.$$

Since i is recurrent,

$$\sum_{l=1}^{\infty} \mathbb{P}_{i,i}^{(l)} = \infty.$$

Thus

$$\sum_{l=1}^{\infty} \mathbb{P}_{j,j}^{(l)} = \infty,$$

so j is recurrent. □

Corollary 3.7 If $i \leftrightarrow j$ and i is transient, then j is transient as well.

Consequently, if a Markov chain is irreducible, then either all states are transient or all states are recurrent. The transient case is impossible when the state space is finite.

Theorem 3.8 If i is recurrent and i does not communicate with state j, then $\mathbb{P}_{i,j} = 0$.

Proof. Suppose for a contradiction that $\mathbb{P}_{i,j} > 0$. Since i and j do not communicate, i is not accessible from j. Then

$$f_{i,i} = \mathbb{P}(\text{the Markov chain visits } i \text{ again} \mid X_0 = i)$$

$$
\begin{aligned}
&= 1 - \mathbb{P}(\text{the Markov chain never visits } i \text{ again} \mid X_0 = i) \\
&\le 1 - \mathbb{P}_{i,j} \\
&< 1.
\end{aligned}
$$

This inequality holds because $\{X_1 = j\} \subseteq \{\text{the Markov chain never visits } i \text{ again}\}$. This contradicts the recurrence of i. Hence $\mathbb{P}_{i,j} = 0$. ☐

Remark 3.9 Under the assumptions of the preceding theorem, we can in fact show that j is not accessible from i; that is, $\mathbb{P}_{i,j}^{(n)} = 0$ for every $n = 1, 2, \ldots$. The key point is that if j were accessible from i, then there would exist some $n > 0$ such that

$$\mathbb{P}(X_n = j, X_{n-1} \neq j, \ldots, X_1 \neq j \mid X_0 = i) > 0.$$

This would contradict the recurrence of i.

Corollary 3.10 If j is accessible from i but i is not accessible from j, then i must be transient.

Lecture 4 — Wednesday, September 17, 2014

Positive Recurrence and Stationary Distributions

Recall that state i is recurrent when $f_{i,i} = 1$ and transient when $f_{i,i} < 1$. Equivalently, i is recurrent when $\mathbb{E}[M_i \mid X_0 = i] = \infty$, where

$$M_i = \sum_{n=1}^{\infty} \mathbb{1}_{\{X_n = i\}},$$

the number of visits to i after time 0. Thus i is transient if and only if $\mathbb{E}[M_i \mid X_0 = i] < \infty$, while i is recurrent if and only if

$$\sum_{n=1}^{\infty} \mathbb{P}_{i,i}^{(n)} = \infty.$$

Similarly, i is transient if and only if

$$\sum_{n=1}^{\infty} \mathbb{P}_{i,i}^{(n)} < \infty.$$

If $i \leftrightarrow j$ and i is recurrent, then j is recurrent. Hence transience and recurrence are class properties. Also, if i is recurrent and $i \not\to j$, then $\mathbb{P}_{i,j} = 0$. Consequently, if j is accessible from i but i is not accessible from j, then i is transient.

Remark 4.1 This shows that once a Markov chain enters a recurrent class, it never leaves that class.

Definition 4.2 Define
$$R_i := \min\{n \geq 1 \mid X_n = i\}$$
to be the first return time to state i.

We have
$$\mathbb{P}(R_i = k \mid X_0 = i) = f_{i,i}^{(k)}.$$
If i is recurrent, then
$$\sum_{k=1}^{\infty} f_{i,i}^{(k)} = f_{i,i} = 1$$
and
$$\mathbb{P}(R_i = \infty \mid X_0 = i) = 0.$$

However, recurrence does *not* imply that
$$\mathbb{E}\left[R_i \mid X_0 = i\right] = \sum_{k=1}^{\infty} k \cdot f_{i,i}^{(k)} < \infty.$$

Definition 4.3 Let i be recurrent. The state i is called **positive recurrent** if
$$\mathbb{E}[R_i \mid X_0 = i] < \infty,$$
and **null recurrent** if
$$\mathbb{E}[R_i \mid X_0 = i] = \infty.$$

Positive and null recurrence are class properties.

Theorem 4.4 Let $i \leftrightarrow j$ and suppose i is positive recurrent. Then j is also positive recurrent.

The proof is beyond the scope of our course.

Definition 4.5 A discrete-time Markov chain is called **ergodic** if it is irreducible, positive recurrent, and aperiodic.

Stationary Distributions

Definition 4.6 A sequence $\{p_i\}_{i\geq 0}$ with $p_i \geq 0$ for all i is called a **stationary distribution** for a discrete-time Markov chain with transition matrix $\mathbf{P}$ if

$$\sum_{i=0}^{\infty} p_i = 1$$

and

$$p_j = \sum_{i=0}^{\infty} p_i \mathbb{P}_{i,j} \quad \text{for all } j.$$

If $\boldsymbol{p} = (p_0, p_1, \dots)^\top$, then these conditions are $\boldsymbol{p} = \mathbf{P}^\top \boldsymbol{p}$ and $\mathbf{1}_{|S|}^\top \boldsymbol{p} = 1$.

This distribution is called stationary because if the chain has initial distribution $\boldsymbol{\alpha}_0 = \boldsymbol{p}$, then

$$\boldsymbol{\alpha}_n = (\mathbf{P}^\top)^n \boldsymbol{\alpha}_0 = (\mathbf{P}^\top)^n \boldsymbol{p} = \boldsymbol{p}.$$

Thus every one-dimensional marginal distribution is $\boldsymbol{p}$. In fact, time homogeneity then makes all finite-dimensional distributions invariant under a common time shift, so the process is stationary in the usual strict sense.

Question 4.7 Does a stationary distribution always exist? And when it does exist, is it unique?

In general, the answer to both questions is no. A stationary distribution does not exist if all states of the chain are either null recurrent or transient. It is not unique if the chain has more than one positive recurrent equivalence class.

Example 4.8 Let

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

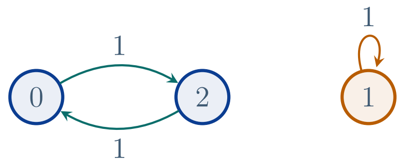


FIGURE 3

Figure 3 shows the two closed classes $\{1\}$ and $\{0, 2\}$. Two stationary distributions are

$$\boldsymbol{p}^{(1)} = (0, 1, 0)^\top \qquad \text{and} \qquad \boldsymbol{p}^{(2)} = \left(\frac{1}{2}, 0, \frac{1}{2}\right)^\top.$$

Hence stationary distributions are not unique. Any convex combination

$$\alpha \boldsymbol{p}^{(1)} + (1-\alpha)\boldsymbol{p}^{(2)} = \left(\frac{1-\alpha}{2}, \alpha, \frac{1-\alpha}{2}\right), \quad \alpha \in [0, 1],$$

is also stationary.

A stationary distribution is sometimes called an **invariant probability**. For a finite state space, the stationary distributions are found by solving the system $\boldsymbol{p} = \mathbf{P}^\top \boldsymbol{p}$ and $\mathbf{1}_N^\top \boldsymbol{p} = 1$. With N states, we have

$$p_j = \sum_{i=0}^{N-1} p_i \mathbb{P}_{i,j}$$

for $j = 0, 1, \ldots, N-1$, and

$$\sum_{j=0}^{N-1} p_j = 1.$$

There are N unknowns $(p_0, \ldots, p_{N-1})$ and $N+1$ equations. If we add the first N equations, we obtain

$$\sum_{j=0}^{N-1} p_j = \sum_{j=0}^{N-1} \sum_{i=0}^{N-1} p_i \mathbb{P}_{i,j} = \sum_{i=0}^{N-1} p_i \sum_{j=0}^{N-1} \mathbb{P}_{i,j} = \sum_{i=0}^{N-1} p_i.$$

Thus one of the balance equations is redundant. To see directly that a stationary distribution exists, start from any initial distribution $\boldsymbol{\alpha}_0$ and average its first n distributions:

$$\overline{\boldsymbol{\alpha}}_n = \frac{1}{n}\sum_{k=0}^{n-1}(\mathbf{P}^\top)^k\boldsymbol{\alpha}_0.$$

These averages lie in the compact probability simplex, so some subsequence converges to a probability vector $\boldsymbol{p}$. Moreover,

$$\mathbf{P}^\top\overline{\boldsymbol{\alpha}}_n - \overline{\boldsymbol{\alpha}}_n = \frac{1}{n}\Big((\mathbf{P}^\top)^n\boldsymbol{\alpha}_0 - \boldsymbol{\alpha}_0\Big) \longrightarrow \mathbf{0}_N.$$

Taking the same subsequential limit gives $\mathbf{P}^\top\boldsymbol{p} = \boldsymbol{p}$. Consequently, every finite-state Markov chain has at least one stationary distribution, although it need not be unique.

Limiting Distributions

We are interested in $\lim_{n\to\infty}\mathbb{P}_{i,j}^{(n)}$ and $\lim_{n\to\infty}\boldsymbol{\alpha}_n$. Periodicity and recurrence play an important role in the limiting behaviour of a Markov chain.

Example 4.9 Using the same 3-state Markov chain as in Example 4.8,

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

Since

$$\mathbf{P}^2 = \mathbf{I},$$

we have

$$\mathbf{P}^{2n} = \mathbf{I} \qquad \text{and} \qquad \mathbf{P}^{2n+1} = \mathbf{P}.$$

Thus $\mathbf{P}^{2n+1} \neq \mathbf{P}^{2n}$, so the limit of $\mathbf{P}^n$ does not exist.

Lecture 5 — Monday, September 22, 2014

Examples show that, in general, a stationary distribution need not exist, and when it exists it need not be unique.

We are interested in limits such as $\lim_{n\to\infty} \mathbb{P}_{i,j}^{(n)}$, $\lim_{n\to\infty} \boldsymbol{\alpha}_n$, and $\lim_{n\to\infty} \mathbb{P}(X_n = j)$.

Example 5.1 Let
$$\mathbf{P} = \begin{pmatrix} 2/4 & 1/4 & 1/4 \\ 1/4 & 2/4 & 1/4 \\ 0 & 1/2 & 1/2 \end{pmatrix}.$$

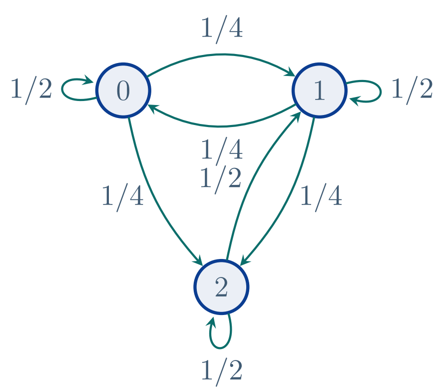


Figure 4

Figure 4 shows that this chain has one class, so it is irreducible and recurrent. Moreover, $d(i) = 1$ for $i = 0, 1, 2$.

For this example,
$$\lim_{n\to\infty} \mathbf{P}^n = \begin{pmatrix} 2/9 & 4/9 & 1/3 \\ 2/9 & 4/9 & 1/3 \\ 2/9 & 4/9 & 1/3 \end{pmatrix}.$$

The limiting row must be a stationary distribution. Solving
$$\boldsymbol{\pi}^\top \mathbf{P} = \boldsymbol{\pi}^\top \qquad \text{and} \qquad \boldsymbol{\pi}^\top \mathbf{1}_3 = 1,$$

gives

$$\boldsymbol{\pi} = \left(\frac{2}{9}, \frac{4}{9}, \frac{1}{3}\right)^\top.$$

To verify the limit using only direct matrix multiplication, put

$$\mathbf{\Pi} = \mathbf{1}_3\boldsymbol{\pi}^\top, \qquad \mathbf{Q} = \mathbf{I} - \mathbf{\Pi}, \qquad \mathbf{N} = \begin{pmatrix} 1/12 & -1/12 & 0 \\ 1/12 & -1/12 & 0 \\ -1/6 & 1/6 & 0 \end{pmatrix}.$$

Direct multiplication gives

$$\mathbf{P} = \mathbf{\Pi} + \frac{1}{4}\mathbf{Q} + \mathbf{N}, \qquad \mathbf{\Pi}\mathbf{Q} = \mathbf{Q}\mathbf{\Pi} = \mathbf{0}, \qquad \mathbf{Q}\mathbf{N} = \mathbf{N}\mathbf{Q} = \mathbf{N}, \qquad \mathbf{N}^2 = \mathbf{0}.$$

An induction now gives

$$\mathbf{P}^n = \mathbf{\Pi} + 4^{-n}\mathbf{Q} + n\,4^{-(n-1)}\mathbf{N}.$$

The last two terms tend to zero, so $\mathbf{P}^n \to \mathbf{\Pi}$, which is the displayed matrix. The limit of $\mathbf{P}^n$ exists, and the limiting matrix has identical rows. This implies that

$$\lim_{n\to\infty} \boldsymbol{\alpha}_n = \left(\lim_{n\to\infty} (\mathbf{P}^\top)^n\right) \boldsymbol{\alpha}_0 = (2/9, 4/9, 1/3)^\top.$$

Thus the limiting distribution of X_n exists and does not depend on the initial distribution.

Example 5.2 Let

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

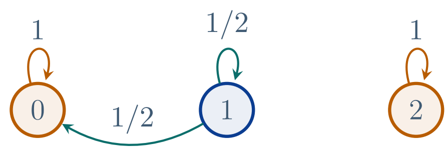


FIGURE 5

Figure 5 shows the classes $\{0\}, \{1\}, \{2\}$. The states $\{0\}$ and $\{2\}$ are recurrent, while $\{1\}$ is transient. The chain is aperiodic. Since $\mathbb{P}_{00} = \mathbb{P}_{22} = 1$, states 0 and 2 are absorbing. We have

$$\lim_{n\to\infty} \mathbb{P}_{00}^{(n)} = \lim_{n\to\infty} \mathbb{P}_{22}^{(n)} = 1,$$

while $\lim_{n\to\infty} \mathbb{P}_{01}^{(n)} = \lim_{n\to\infty} \mathbb{P}_{02}^{(n)} = \lim_{n\to\infty} \mathbb{P}_{20}^{(n)} = \lim_{n\to\infty} \mathbb{P}_{21}^{(n)} = 0$ and $\mathbb{P}_{11}^{(n)} \to 0$. Thus

$$\lim_{n\to\infty} \mathbf{P}^n = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The limit of $\mathbf{P}^n$ exists, but the rows are not identical. As a result, the limiting distribution depends on the initial state. The column corresponding to state 1 is the zero vector in the limit, so $((\mathbf{P}^\top)^n \boldsymbol{\alpha}_0)_1 \to 0$ as $n \to \infty$ for every initial distribution $\boldsymbol{\alpha}_0$. Thus, no matter what the initial state or distribution is, the limiting distribution has no mass at state 1.

Proposition 5.3 Let j be a transient state of a Markov chain. Then for any state i,

$$\lim_{n\to\infty} \mathbb{P}_{i,j}^{(n)} = 0.$$

Proof.

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathbb{P}_{i,j}^{(n)} &= \sum_{n=1}^{\infty} \sum_{k=1}^{n} f_{i,j}^{(k)} \mathbb{P}_{j,j}^{(n-k)} \\
&= \sum_{k=1}^{\infty} f_{i,j}^{(k)} \sum_{n=k}^{\infty} \mathbb{P}_{j,j}^{(n-k)} \\
&= \left(\sum_{k=1}^{\infty} f_{i,j}^{(k)} \right) \left(\sum_{m=0}^{\infty} \mathbb{P}_{j,j}^{(m)} \right) \\
&= f_{i,j} \left(1 + \sum_{m=1}^{\infty} \mathbb{P}_{j,j}^{(m)} \right) < \infty,
\end{aligned}$$

since j is transient. Therefore $\mathbb{P}_{i,j}^{(n)} \to 0$. □

Theorem 5.4 (Basic limit theorem) For an irreducible, recurrent, aperiodic Markov chain, $\lim_{n\to\infty} \mathbb{P}_{i,j}^{(n)}$ exists and is independent of the initial state i. Furthermore,

$$\pi_j := \lim_{n\to\infty} \mathbb{P}_{i,j}^{(n)} = \frac{1}{m_j}, \qquad m_j := \mathbb{E}\left[R_j \mid X_0 = j\right],$$

with the convention $1/\infty = 0$. If the chain is positive recurrent, then $\{\pi_j\}_{j\geq 0}$ is the unique stationary distribution.

Remark 5.5 Suppose that, for some initial state i, all limits $\pi_j = \lim_{n\to\infty} \mathbb{P}_{ij}^{(n)}$ exist and satisfy $\sum_j \pi_j = 1$. Then $\boldsymbol{\pi}$ is a stationary distribution. Indeed, pointwise convergence of probability mass functions to another probability mass function implies convergence in ℓ^1. Therefore the Chapman–Kolmogorov equation

$$\mathbb{P}_{ij}^{(n)} = \sum_{k\in S} \mathbb{P}_{ik}^{(n-1)} \mathbb{P}_{kj}.$$

allows the limit to pass through the sum, giving

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}_{ij}^{(n)} &= \sum_{k=0}^{\infty} \lim_{n\to\infty} \mathbb{P}_{ik}^{(n-1)} \mathbb{P}_{kj} \\ &= \sum_{k=0}^{\infty} \pi_k \mathbb{P}_{kj}.\end{aligned}$$

Thus $\boldsymbol{\pi} = \mathbf{P}^\top \boldsymbol{\pi}$. Without the total-mass condition, probability can escape to infinity and the pointwise limits need not constitute a distribution.

Lecture 6 — Wednesday, September 24, 2014

The examples above show that, for an irreducible aperiodic discrete-time Markov chain, $\lim_{n\to\infty} \mathbf{P}^n$ can exist and have identical rows. In that case the limiting distribution does not depend on the initial state or initial distribution. For aperiodic chains with more than one positive recurrent class, $\lim_{n\to\infty} \mathbf{P}^n$ can exist but need not have identical rows, so the limiting distribution can depend on the initial state. If j is transient, then $\lim_{n\to\infty} \mathbb{P}_{ij}^{(n)} = 0$ for every i.

The basic limit theorem says that if a Markov chain is irreducible, recurrent, and aperiodic, then $\lim_{n\to\infty} \mathbb{P}_{ij}^{(n)}$ exists and is independent of i. Furthermore,

$$\pi_j = \lim_{n\to\infty} \mathbb{P}_{ij}^{(n)} = \frac{1}{m_j}.$$

For an irreducible positive recurrent Markov chain, the stationary distribution $\boldsymbol{\pi}$ also represents the fraction of time the chain spends in different states in the long run, regardless of the initial state i. Aperiodicity is not needed for this occupation-time statement. The fraction of time spent in state j up to time $n-1$ is

$$\gamma_{i,j}^{(n)} := \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_{\{X_k=j\}}.$$

The cycle argument uses the standard restart principle: after a hitting or return time, the future chain has the same transition law as a fresh chain from the state reached and is independent of the preceding path. Its complete justification requires the strong Markov property, which is beyond the scope of this course. Accepting this principle, the cycles between successive visits to j are independent and identically distributed. The initial segment before the first visit is almost surely finite and does not affect the limiting occupation fraction. The strong law of large numbers applied to the return cycles gives

$$\gamma_{i,j}^{(n)} \longrightarrow \frac{1}{\mathbb{E}[R_j \mid X_0 = j]} = \frac{1}{m_j} = \pi_j \quad \text{almost surely.}$$

Moreover,

$$\mathbb{E}[\gamma_{i,j}^{(n)} \mid X_0 = i] = \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{P}_{ij}^{(k)} \longrightarrow \pi_j.$$

Thus the Cesàro averages of the transition probabilities converge even when the chain is periodic. When the chain is also aperiodic, the basic limit theorem strengthens this conclusion to $\mathbb{P}_{ij}^{(n)} \to \pi_j$.

Remark 6.1

1. For an irreducible positive recurrent periodic Markov chain, $\lim_{n\to\infty} \mathbb{P}_{ij}^{(n)}$ does not exist in general, but
$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{P}_{ij}^{(k)}$$
always exists, and the occupation fraction $\gamma_{i,j}^{(n)}$ converges almost surely.
2. In the periodic case, there is an extension of the basic limit theorem:
$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{P}_{ij}^{(k)} = \frac{1}{m_j}.$$

For example, if

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

then $d = 2$ and $\lim_{n\to\infty} \mathbf{P}^{2n} = \mathbf{I}$, so in particular $\mathbb{P}_{00}^{(2n)} \to 1$, while $1/M_{00} = 1/2$.

Theorem 6.2 If state i is null recurrent, then $\lim_{n\to\infty} \mathbb{P}_{ij}^{(n)} = 0$ for any state j.

Proof. If $i = j$, the renewal theorem gives $\mathbb{P}_{ii}^{(nd)} \to d/m_i = 0$ along the possible return-time lattice, where d is the period; for n outside that lattice the return probability is 0. Hence $\mathbb{P}_{ii}^{(n)} \to 0$. If $i \neq j$, there are two cases:

1. If i and j are not in the same class, then since i is recurrent, j is not accessible from i. Hence $\mathbb{P}_{ij}^{(n)} = 0$ and $\lim_{n\to\infty} \mathbb{P}_{ij}^{(n)} = 0$.
2. If i and j are in the same class, then j is also null recurrent. Recall that

$$\mathbb{P}_{ij}^{(n)} = \sum_{k=1}^{n} f_{ij}^{(k)} \mathbb{P}_{jj}^{(n-k)}.$$

Fix $\varepsilon > 0$ and choose K so that $\sum_{k>K} f_{ij}^{(k)} < \varepsilon$. Then

$$\mathbb{P}_{ij}^{(n)} \leq \sum_{k=1}^{K} f_{ij}^{(k)} \mathbb{P}_{jj}^{(n-k)} + \varepsilon.$$

The finite sum tends to 0 because $\mathbb{P}_{jj}^{(n-k)} \to 0$ for every fixed k. Hence $\limsup_{n\to\infty} \mathbb{P}_{ij}^{(n)} \leq \varepsilon$, and letting $\varepsilon \downarrow 0$ proves the claim.

□

Remark 6.3 The second case above also shows that $\lim_{n\to\infty} \mathbb{P}_{ij}^{(n)} = 0$ for all null recurrent j.

Theorem 6.4 In a finite-state Markov chain, there are no null recurrent states.

Proof. Suppose, for a contradiction, that i is a null recurrent state. By Remark 6.3, $\lim_{n\to\infty} \mathbb{P}_{ij}^{(n)} = 0$ for every $j \in S$. Therefore

$$1 = \lim_{n\to\infty} \sum_{j\in S} \mathbb{P}_{ij}^{(n)} = \sum_{j\in S} \lim_{n\to\infty} \mathbb{P}_{ij}^{(n)} = \sum_{j\in S} 0 = 0,$$

which is a contradiction. $\square$

Theorem 6.5 An irreducible Markov chain is positive recurrent if and only if a stationary distribution exists.

Proof. *Forward implication.* Assume that the chain is positive recurrent and fix a state i. Let R_i be its first return time and define the expected occupation measure of one return cycle by

$$\nu_j = \mathbb{E}\left[\sum_{k=0}^{R_i-1} \mathbb{1}_{\{X_k=j\}} \mid X_0 = i\right], \quad j \in S.$$

Since $\sum_{j\in S} \nu_j = \mathbb{E}[R_i \mid X_0 = i] < \infty$, the vector

$$\pi_j = \frac{\nu_j}{\mathbb{E}[R_i \mid X_0 = i]}$$

is a probability distribution. Moreover,

$$\begin{aligned}\sum_{\ell\in S} \nu_\ell \mathbb{P}_{\ell j} &= \mathbb{E}\left[\sum_{k=0}^{R_i-1} \mathbb{P}_{X_k j} \mid X_0 = i\right] \\ &= \mathbb{E}\left[\sum_{k=1}^{R_i} \mathbb{1}_{\{X_k=j\}} \mid X_0 = i\right] = \nu_j.\end{aligned}$$

The final equality holds because the cycle begins and ends at i. Thus $\mathbf{P}^\top \boldsymbol{\nu} = \boldsymbol{\nu}$, and normalizing gives $\mathbf{P}^\top \boldsymbol{\pi} = \boldsymbol{\pi}$.

Lecture 7 — Monday, September 29, 2014

Reverse implication. Suppose that a stationary distribution $\boldsymbol{\pi}$ exists, but the chain is null recurrent or transient. Null recurrence and transience imply that $\lim_{n\to\infty} \mathbb{P}_{ij}^{(n)} = 0$ for every $i, j \in S$. However, for a stationary distribution π,

$$\pi_j = \sum_{i\in S} \pi_i \mathbb{P}_{ij}^{(n)} \quad \text{for all } j \text{ and } n = 1, 2, 3, \ldots.$$

Fix j and $\varepsilon > 0$. Choose a finite set $F \subseteq S$ such that $\sum_{i\notin F} \pi_i < \varepsilon$. Since F is finite and $\mathbb{P}_{ij}^{(n)} \to 0$ for every $i \in F$,

$$\sum_{i\in F} \pi_i \mathbb{P}_{ij}^{(n)} \longrightarrow 0.$$

For every n, the remaining terms satisfy

$$0 \le \sum_{i\notin F} \pi_i \mathbb{P}_{ij}^{(n)} \le \sum_{i\notin F} \pi_i < \varepsilon.$$

Since ε was arbitrary, $\pi_j = 0$. Since this conclusion holds for every j, we have $\boldsymbol{\pi} = (0, 0, \ldots)$, which is not a distribution. This contradiction proves the converse direction. □

Theorem 7.1 Let $\{X_n\}_{n\ge 0}$ be an irreducible, positive recurrent Markov chain. Fix $i \in S$. For $j \in S$, define

$$\nu_j = \mathbb{E}\left[\sum_{k=0}^{R_i - 1} \mathbb{1}_{\{X_k = j\}} \mid X_0 = i\right]$$

to be the expected number of visits to j before returning to i, given $X_0 = i$. If $\boldsymbol{\nu} = (\nu_j)_{j\in S}$, then

$$\boldsymbol{\pi} = \frac{\boldsymbol{\nu}}{\sum_{j\in S} \nu_j}$$

is the stationary distribution of $\{X_n\}$.

To see why this formula holds, set $R_i(0) = 0$ and define $R_i(k)$ to be the time of the kth revisit to state i, so $R_i(1) = R_i$. Let $Y_j(k)$ be the number of visits to state j between the $(k-1)$st and kth visits to i. Then

$$Y_j(k) = \sum_{n=R_i(k-1)}^{R_i(k)-1} \mathbb{1}_{\{X_n = j\}}.$$

By the restart principle described above, the random variables $Y_j(k)$ are independent and identically distributed. By the strong law of large numbers, given $X_0 = i$,

$$\frac{1}{m}\sum_{k=1}^{m} Y_j(k) \to \mathbb{E}\left[Y_j(1) \mid X_0 = i\right] = \nu_j$$

as $m \to \infty$. On the other hand,

$$\begin{aligned}\lim_{m\to\infty} \frac{1}{m}\sum_{k=1}^{m} Y_j(k) &= \lim_{m\to\infty} \frac{\sum_{n=0}^{R_i(m)-1} \mathbb{1}_{\{X_n=j\}}}{R_i(m)} \cdot \frac{R_i(m)}{m} \\ &= \text{long-run proportion of time spent in } j \cdot m_i.\end{aligned}$$

Thus ν_j is the long-run proportion of time spent in j multiplied by the positive constant m_i. Since the long-run proportions form the stationary distribution, normalizing $\boldsymbol{\nu}$ gives the stationary distribution. Notice that

$$\sum_{j\in S} \nu_j = m_i$$

in this setting, and

$$\frac{\nu_j}{\nu_i} = \frac{\pi_j}{\pi_i}.$$

Here is a nice theorem which we note in passing:

Theorem 7.2 (Markov chain ergodic theorem) For a positive recurrent, irreducible Markov chain with stationary distribution $\boldsymbol{\pi}$, consider a **reward function** $f : S \to \mathbb{R}$. If

$$\sum_{j\in S} \pi_j |f(j)| < \infty,$$

then

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f(X_k) = \sum_{i\in S} \pi_i f(i) \quad \text{almost surely.}$$

That is, the time average reward converges to the stationary average reward.

Reducible Markov Chains

The canonical decomposition expresses the state space S of a Markov chain as

$$S = T \cup C_1 \cup C_2 \cup \ldots,$$

where T consists of all transient states, not necessarily forming a single class, and the C_i are closed disjoint classes of recurrent states.

Definition 7.3 A set of states C is **closed** if $\mathbb{P}_{ij} = 0$ for all $i \in C$ and $j \in S \setminus C$.

Thus, if $i \in C$ and $j \in S \setminus C$, then $\mathbb{P}(\text{ever visits } j \mid X_0 = i) = 0$. After ordering the transient states first and then the recurrent classes, the transition matrix has the canonical block form

$$\mathbf{P} = \begin{pmatrix} \mathbf{Q} & \mathbf{R}_1 & \mathbf{R}_2 & \cdots \\ \mathbf{0} & \mathbf{P}_1 & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{P}_2 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

where $\mathbf{Q}$ describes transitions among transient states, the $\mathbf{R}_r$ describe transitions from transient states into recurrent classes, and $\mathbf{P}_r$ is the transition matrix on recurrent class C_r. The restriction to the union of recurrent classes is block diagonal. If the chain starts in a recurrent class C, we can restrict attention to the chain on C.

Example 7.4 Consider a Markov chain with

$$\mathbf{P} = \begin{pmatrix} p & 0 & 1-p & 0 \\ 0 & r & 0 & 1-r \\ q & 0 & 1-q & 0 \\ 0 & s & 0 & 1-s \end{pmatrix}$$

where $0 < p, q, r, s < 1$. These restrictions make each of the two classes below irreducible and aperiodic. This chain is not irreducible. There are two classes, $\{0, 2\}$ and $\{1, 3\}$, both positive recurrent. Rearranging the terms gives

$$\mathbf{P}^* = \begin{pmatrix} p & 1-p & 0 & 0 \\ q & 1-q & 0 & 0 \\ 0 & 0 & r & 1-r \\ 0 & 0 & s & 1-s \end{pmatrix} = \begin{pmatrix} \mathbf{P}_1 & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 2} & \mathbf{P}_2 \end{pmatrix}$$

where $\mathbf{P}_1$ is the transition probability matrix for the first class and $\mathbf{P}_2$ is the transition probability matrix for the second class. Then

$$(\mathbf{P}^*)^n = \begin{pmatrix} \mathbf{P}_1^n & 0 \\ 0 & \mathbf{P}_2^n \end{pmatrix}.$$

For the class $\{0, 2\}$,

$$\mathbf{P}_1 = \begin{pmatrix} p & 1-p \\ q & 1-q \end{pmatrix}.$$

Its stationary distribution (π_0, π_2) satisfies

$$(\pi_0, \pi_2)\mathbf{P}_1 = (\pi_0, \pi_2) \qquad \text{and} \qquad \pi_0 + \pi_2 = 1.$$

Hence

$$\pi_0(1-p) = \pi_2 q \qquad \text{and} \qquad (\pi_0, \pi_2) = \left(\frac{q}{q+1-p}, \frac{1-p}{q+1-p}\right).$$

Therefore

$$\lim_{n\to\infty} \mathbf{P}_1^n = \begin{pmatrix} \dfrac{q}{q+1-p} & \dfrac{1-p}{q+1-p} \\ \dfrac{q}{q+1-p} & \dfrac{1-p}{q+1-p} \end{pmatrix}.$$

For the class $\{1, 3\}$,

$$\mathbf{P}_2 = \begin{pmatrix} r & 1-r \\ s & 1-s \end{pmatrix}.$$

Its stationary distribution (π_1, π_3) satisfies

$$(\pi_1, \pi_3)\mathbf{P}_2 = (\pi_1, \pi_3) \qquad \text{and} \qquad \pi_1 + \pi_3 = 1,$$

so

$$\pi_1(1-r) = \pi_3 s \qquad \text{and} \qquad (\pi_1, \pi_3) = \left(\frac{s}{s+1-r}, \frac{1-r}{s+1-r}\right).$$

Thus

$$\lim_{n\to\infty} \mathbf{P}_2^n = \begin{pmatrix} \dfrac{s}{s+1-r} & \dfrac{1-r}{s+1-r} \\ \dfrac{s}{s+1-r} & \dfrac{1-r}{s+1-r} \end{pmatrix}.$$

Returning to the original state order $(0,1,2,3)$, we obtain

$$\lim_{n\to\infty} \mathbf{P}^n = \begin{pmatrix} \frac{q}{q+1-p} & 0 & \frac{1-p}{q+1-p} & 0 \\ 0 & \frac{s}{s+1-r} & 0 & \frac{1-r}{s+1-r} \\ \frac{q}{q+1-p} & 0 & \frac{1-p}{q+1-p} & 0 \\ 0 & \frac{s}{s+1-r} & 0 & \frac{1-r}{s+1-r} \end{pmatrix}.$$

The stationary distributions of the full reducible chain are not unique. They are exactly the convex combinations of the stationary distributions for the two classes:

$$\boldsymbol{\pi}^{(\lambda)} = \left(\lambda\frac{q}{q+1-p}, (1-\lambda)\frac{s}{s+1-r}, \lambda\frac{1-p}{q+1-p}, (1-\lambda)\frac{1-r}{s+1-r}\right), \quad 0 \le \lambda \le 1.$$

Lecture 8 — Wednesday, October 01, 2014

Recall that irreducible Markov chain is positive recurrent if and only if a stationary distribution exists. For an irreducible, positive recurrent Markov chain, the long-run proportion of time spent in each state forms a stationary distribution. Thus, if $\{X_n\}$ is irreducible and positive recurrent, define

$$\nu_j = \mathbb{E}\left[\sum_{k=0}^{R_i-1} \mathbb{1}_{\{X_k=j\}} \mid X_0 = i\right].$$

Then

$$\pi_j = \frac{\nu_j}{\sum_{\ell\in S}\nu_\ell}$$

defines the stationary distribution of $\{X_n\}$.

The long-run behaviour is summarized by

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} f(X_i) = \pi\cdot f^T.$$

For reducible Markov chains, the canonical decomposition is $S = T \cup C_1 \cup C_2 \cup \dots$.

For a fixed initial state i, any pointwise limit that retains total mass 1 must be a stationary distribution supported on reachable positive recurrent closed classes. Aperiodicity of every reachable positive recurrent class is a convenient sufficient condition for convergence, but it is not necessary for specially balanced initial or entrance distributions. A stationary distribution exists if and only if the chain has at least one positive recurrent class, and it is unique if and only if it has exactly one positive recurrent class. Every stationary distribution assigns zero mass to transient and null recurrent states. In the irreducible case, positive recurrence and aperiodicity ensure that the limiting distribution exists and agrees with the unique stationary distribution. The occupation-time limit agrees with that stationary distribution even without aperiodicity.

Simple Random Walk

Definition 8.1 Let $\{X_n\}_{n\geq 1}$ be independent and identically distributed, with

$$\mathbb{P}(X_i = 1) = p = 1 - q = 1 - \mathbb{P}(X_i = -1).$$

Define the **simple random walk** $\{S_n\}_{n\geq 0}$ by $S_0 = 0$ and $S_n = X_1 + \cdots + X_n$.

The simple random walk is a Markov chain because

$$\begin{aligned}
&\mathbb{P}(S_{n+1} = s_{n+1} \mid S_n = s_n, \ldots, S_0 = 0) \\
&\qquad = \mathbb{P}(X_{n+1} = s_{n+1} - s_n \mid S_n = s_n, \ldots, S_0 = 0) \\
&\qquad = \mathbb{P}(X_{n+1} = s_{n+1} - s_n) \\
&\qquad = \mathbb{P}(S_{n+1} = s_{n+1} \mid S_n = s_n).
\end{aligned}$$

If $p = 1$ or $q = 1$, the chain is deterministic, all states are transient, and each state forms its own equivalence class. If $p \in (0, 1)$, the chain is irreducible with period $d = 2$.

Proposition 8.2 The simple random walk is transient if $p \neq q$, and null recurrent if $p = q = 1/2$.

Proof. If $p \neq q$, assume without loss of generality that $p > q$. By the strong law of large numbers, $S_n/n \to p - q > 0$, so $S_n \to \infty$ almost surely. Then there is a last visit to state 0, so state 0 is transient, and the chain is transient. If $p = 1/2$, we show

$$\sum_{n=0}^{\infty} \mathbb{P}_{00}^{(2n)} = \infty$$

to prove recurrence. Since $d = 2$, $\mathbb{P}_{00}^{(2n+1)} = 0$ for all $n = 0, 1, \dots$. Also,

$$\mathbb{P}_{00}^{(2n)} = \mathbb{P}(n \text{ steps left, } n \text{ steps right}) = \binom{2n}{n}\left(\frac{1}{2}\right)^n\left(\frac{1}{2}\right)^n = \binom{2n}{n}\left(\frac{1}{4}\right)^n$$

Using Stirling's formula $n! \sim \sqrt{2\pi n}e^{-n}n^n$,

$$\binom{2n}{n} = \frac{(2n)!}{(n!)^2} \sim \frac{\sqrt{4\pi n}e^{-2n}(2n)^{2n}}{2\pi n e^{-2n}n^{2n}} = \frac{2\sqrt{\pi n}2^{2n}n^{2n}}{2\pi n n^{2n}} = \frac{2^{2n}}{\sqrt{\pi n}}$$

Then

$$\mathbb{P}_{00}^{(2n)} \sim \left(\frac{2^{2n}}{\sqrt{\pi n}}\right)\left(\frac{1}{4}\right)^n = \frac{1}{\sqrt{\pi n}}.$$

Hence comparison with the divergent series $\sum_{n\geq 1} n^{-1/2}$ gives

$$\sum_{n=0}^{\infty} \mathbb{P}_{00}^{(2n)} = \infty.$$

Therefore state 0 is recurrent.

On the other hand, solve the system of equations $\mathbf{P}^\top \boldsymbol{\pi} = \boldsymbol{\pi}$. The ith component satisfies

$$\pi_i = \frac{1}{2}\pi_{i-1} + \frac{1}{2}\pi_{i+1},$$

so $\pi_{i+1} - \pi_i = \pi_i - \pi_{i-1}$. Since this holds for all $i \in \mathbb{Z}$, the sequence $\{\pi_i\}$ is arithmetic, and $\pi_i = \pi_0 + id$, where $d = \pi_1 - \pi_0$. The only possible nonnegative arithmetic sequence on all of $\mathbb{Z}$ is constant, so $\pi_i = \pi_0$. If $\pi_0 = 0$, then all $\pi_i = 0$ and

$$\sum_{i\in S} \pi_i \neq 1.$$

If $\pi_0 > 0$, then

$$\sum_{i\in S} \pi_i = \infty \neq 1.$$

Therefore no stationary distribution exists. By Theorem 6.5, the chain cannot be positive recurrent. Since it is recurrent, it must be null recurrent. □

Another useful calculation for the simple random walk is to find

$$f_{00} = \mathbb{P}(\text{the Markov chain visits 0 again} \mid X_0 = 0)$$

and, more generally,

$$f_{i0} = \mathbb{P}(\text{visit } 0 \mid X_0 = i).$$

Proposition 8.3 For simple random walk,

$$f_{00} = 2\min\{p, 1-p\}$$

and, for $i \geq 1$,

$$f_{i,0} = \begin{cases} 1, & p \leq \frac{1}{2}, \\ \left(\dfrac{1-p}{p}\right)^i, & p > \frac{1}{2}. \end{cases}$$

Proof. Write $q = 1-p$. First suppose $p > q$, and let $h_i = f_{i,0}$ for $i \geq 0$, with $h_0 = 1$ because this is a hitting probability rather than a return probability. A first-step analysis gives

$$h_i = ph_{i+1} + qh_{i-1}, \qquad i \geq 1.$$

For $K > i$, stop the walk upon first reaching 0 or K. Solving the finite gambler's-ruin recurrence gives

$$\mathbb{P}_i(T_0 < T_K) = \frac{(q/p)^i - (q/p)^K}{1-(q/p)^K}.$$

As $K \to \infty$, the events $\{T_0 < T_K\}$ increase to $\{T_0 < \infty\}$, and therefore

$$f_{i,0} = h_i = \left(\frac{q}{p}\right)^i, \qquad i \geq 1.$$

Starting from 0, a first step to -1 is followed by an almost-sure return to 0, whereas a first step to 1 is followed by a return with probability q/p. Hence

$$f_{00} = q + p\frac{q}{p} = 2q.$$

If $p = q = 1/2$, recurrence gives $f_{i,0} = f_{00} = 1$. If $p < q$, reflection of the walk interchanges p and q, so $f_{i,0} = 1$ for $i \geq 1$ and $f_{00} = 2p$. Hence

$$f_{00} = 2\min\{p, 1-p\}.$$

$\square$

Lecture 9 — Monday, October 06, 2014

For any fixed i, pointwise limits $\lim_{n\to\infty} \mathbb{P}_{ij}^{(n)}$, when they exist, form a distribution only if no probability mass escapes to infinity. In reducible chains the limiting law can be a mixture over several reachable positive recurrent classes, with weights determined by the corresponding entrance probabilities.

For the simple random walk, the chain is transient if $p \neq q$ and null recurrent if $p = q$. We also found that

$$f_{00} = 2\min\{p, 1-p\}.$$

G/M/1 Queues

Consider a discrete-time Markov chain with transition probability matrix

$$\mathbf{P} = \begin{pmatrix} 1-b_0 & b_0 & 0 & 0 & 0 & \dots \\ 1-(b_0+b_1) & b_1 & b_0 & 0 & 0 & \dots \\ 1-(b_0+b_1+b_2) & b_2 & b_1 & b_0 & 0 & \dots \\ 1-(b_0+b_1+b_2+b_3) & b_3 & b_2 & b_1 & b_0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

where $\boldsymbol{b} = (b_0, b_1, \dots)$ is a probability distribution.

In the notation $G/M/1$, G indicates a general **interarrival time distribution**, M indicates exponentially distributed **service times**, and 1 indicates one server. We are interested in the queue length at arrival times. Let X_n be the number of customers present upon the nth arrival, not including the arriving customer. The distribution $\boldsymbol{b}$ describes the number of potential service completions between two successive arrivals, meaning the number that would occur if the server never became idle:

$$b_k = \mathbb{P}(\text{there are } k \text{ potential service completions between two arrivals}).$$

Assume $b_0 > 0$ and $b_0 + b_1 < 1$; these conditions imply that the chain is irreducible and aperiodic.

We want to find a condition for this chain to be positive recurrent.

Find a stationary distribution $\boldsymbol{\pi} = (\pi_0, \pi_1, \dots)^\top$ satisfying $\boldsymbol{\pi} = \mathbf{P}^\top \boldsymbol{\pi}$:

$$\pi_0 = \left(1 - \sum_{j=0}^{0} b_j\right) \pi_0 + \left(1 - \sum_{j=0}^{1} b_j\right) \pi_1 + \cdots = \sum_{i=0}^{\infty} \pi_i \left(1 - \sum_{j=0}^{i} b_j\right) \tag{9.1}$$

$$\pi_1 = b_0\pi_0 + b_1\pi_1 + b_2\pi_2 + \dots \tag{9.2}$$

$$\pi_2 = b_0\pi_1 + b_1\pi_2 + \dots \tag{9.3}$$

In general,

$$\pi_k = \sum_{i=0}^{\infty} b_i \pi_{k-1+i}$$

with

$$\sum_{i=0}^{\infty} \pi_i = 1.$$

Assume the solution has the form $\pi_k = r_0^k(1 - r_0)$ for $k = 0, 1, 2, \dots$ and $r_0 \in (0, 1)$. We want to verify this form in (9.1), (9.2), and (9.3), and then determine r_0. For $k \geq 1$,

$$\pi_k = \sum_{i=0}^{\infty} b_i r_0^{k-1+i}(1 - r_0) = \sum_{i=0}^{\infty} b_i r_0^{k-1+i}(1 - r_0) \implies r_0 = \sum_{i=0}^{\infty} b_i r_0^i.$$

Let B be the random variable with probability mass function $\mathbb{P}(B = k) = b_k$. Its generating function is

$$B(z) = \mathbb{E}\left[z^B\right] = \sum_{i=0}^{\infty} z^i b_i.$$

Thus r_0 is a solution of $z = B(z)$, and we want this equation to have a solution in $(0, 1)$. Because r_0 is a solution of $z = B(z)$, it is an intersection of $y = z$ and $y = B(z)$.

The generating function B has the following properties. First, B is continuous. Second,

$$B(0) = \sum_{i=0}^{\infty} z^i b_i\big|_{z=0} = b_0 > 0 \qquad \text{and} \qquad B(1) = \sum_{i=0}^{\infty} b_i = 1,$$

while

$$B'(z) = \frac{\mathrm{d}}{\mathrm{d}z}\left(\sum_{i=0}^{\infty} z^i b_i\right) = \sum_{i=1}^{\infty} i z^{i-1} b_i > 0$$

for $z \in (0, 1)$ and

$$B'(1) = \sum_{i=1}^{\infty} i b_i = \mathbb{E}\left[B\right],$$

and

$$B''(z) = \sum_{i=2}^{\infty} i(i-1)z^{i-2}b_i > 0$$

for $z \in (0,1)$, so B is convex.

If $B'(1) = \mathbb{E}[B] > 1$, then there is a unique solution in $(0,1)$. If $B'(1) \leq 1$, then there is no solution r_0 in $(0,1)$.

We want to verify that the proposed solution satisfies (9.1):

$$\pi_0 \stackrel{?}{=} \sum_{i=0}^{\infty} \pi_i \left(1 - \sum_{j=0}^{i} b_j\right).$$

To this end, we have

$$\begin{aligned}
\text{RHS} &= \sum_{i=0}^{\infty} \pi_i - \sum_{i=0}^{\infty}\sum_{j=0}^{i} \pi_i b_j \\
&= 1 - \sum_{j=0}^{\infty}\sum_{i=j}^{\infty} \pi_i b_j \\
&= 1 - \sum_{j=0}^{\infty} b_j \sum_{i=j}^{\infty} r_0^i(1-r_0) \\
&= 1 - \sum_{j=0}^{\infty} b_j r_0^j \\
&= 1 - B(r_0) \\
&= 1 - r_0 \\
&= \pi_0 \\
&= \text{LHS}.
\end{aligned}$$

Thus we can conclude that with the conditions

$$b_0 > 0, b_0 + b_1 < 1, \mathbb{E}\left[B\right] > 1,$$

there exists a stationary distribution. Hence the chain is positive recurrent. In general, $r_0 = B(r_0)$ does not have an explicit solution and must be found numerically. For example, starting with $r_0^{(0)} = 0$, the iteration $r_0^{(n+1)} = B(r_0^{(n)})$ increases to the smallest fixed point, which is the desired root in $(0,1)$. As a special case, suppose that $b_k = \alpha^k(1-\alpha)$ for $k = 0, 1, 2, \ldots$. Then

$$B(z) = \sum_{k=0}^{\infty} \alpha^k (1-\alpha) z^k = (1-\alpha) \sum_{k=0}^{\infty} (\alpha z)^k = \frac{1-\alpha}{1-\alpha z}$$

for $|\alpha z| < 1$. The equation $z = B(z)$ becomes

$$z = \frac{1-\alpha}{1-\alpha z} \implies \alpha z^2 - z + (1-\alpha) = 0 \implies z_1 = 1, z_2 = \frac{1-\alpha}{\alpha},$$

When $\alpha > 1/2$, the condition $B'(1) > 1$ gives $z_2 \in (0,1)$ and hence $r_0 = (1-\alpha)/\alpha$. If $\mathbb{E}[B] \leq 1$, then $\mathbb{E}[B] < 1$ implies that the chain is transient, while $\mathbb{E}[B] = 1$ implies that the chain is null recurrent.

Lecture 10 — Wednesday, October 08, 2014

For the $G/M/1$ queue, let B be the random variable with probability mass function $\mathbb{P}(B = k) = b_k$, representing the number of services completed between two arrivals. If $\mathbb{E}[B] > 1$, then the chain is positive recurrent. The stationary distribution is

$$\pi_k = r_0^k (1 - r_0), \quad k = 0, 1, 2, \ldots,$$

where r_0 is the unique solution of $z = B(z)$ in $(0,1)$. If $\mathbb{E}[B] = 1$, then the chain is null recurrent, while if $\mathbb{E}[B] < 1$, then the chain is transient.

Absorption

Example 10.1 Consider a Markov chain with transition probability matrix

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 \\ \alpha & \beta & \gamma \\ 0 & 0 & 1 \end{pmatrix}$$

where $\alpha, \beta, \gamma > 0$ and $\alpha + \beta + \gamma = 1$. State 1 is transient, while states 0 and 2 are absorbing. Let

$$T = \min\{n \geq 0 \mid X_n = 0 \text{ or } X_n = 2\},$$

the time until absorption. We want to find the absorption probability to state 0,

$$u = \mathbb{P}(X_T = 0 \mid X_0 = 1),$$

the expected time to absorption, $v = \mathbb{E}[T \mid X_0 = 1]$, and the distribution of T.

A first-step analysis gives

$$\begin{aligned}
u &= \mathbb{P}(X_T = 0 \mid X_0 = 1) \\
&= \mathbb{P}(X_T = 0 \mid X_1 = 0, X_0 = 1)\mathbb{P}(X_1 = 0 \mid X_0 = 1) \\
&\qquad + \mathbb{P}(X_T = 0 \mid X_1 = 1, X_0 = 1)\mathbb{P}(X_1 = 1 \mid X_0 = 1) \\
&\qquad + \mathbb{P}(X_T = 0 \mid X_1 = 2, X_0 = 1)\mathbb{P}(X_1 = 2 \mid X_0 = 1) \\
&= \alpha \cdot 1 + \beta \cdot u + \gamma \cdot 0 \\
&= \alpha + \beta u.
\end{aligned}$$

Thus

$$u = \frac{\alpha}{1-\beta}.$$

For the expected time to absorption,

$$\begin{aligned}
v &= \mathbb{E}\left[T \mid X_0 = 1\right] \\
&= \sum_{i \geq 0} \mathbb{E}\left[T \mid X_1 = i, X_0 = 1\right] \cdot \mathbb{P}(X_1 = i \mid X_0 = 1) \\
&= 1 \cdot \alpha + \mathbb{E}\left[T \mid X_1 = 1, X_0 = 1\right] \cdot \beta + 1 \cdot \gamma \\
&= \alpha + \gamma + (v+1)\beta.
\end{aligned}$$

Therefore

$$v = \frac{1}{1-\beta}.$$

For the distribution of T,

$$\begin{aligned}
&\mathbb{P}(T = k \mid X_0 = 1) \\
&\quad = \mathbb{P}(X_k = 0 \text{ or } 2, X_{k-1} = 1, \ldots, X_1 = 1 \mid X_0 = 1) \\
&\quad = \mathbb{P}(X_k = 0 \text{ or } 2 \mid X_{k-1} = 1, \ldots, X_0 = 1) \\
&\qquad \cdot \mathbb{P}(X_{k-1} = 1 \mid X_{k-2} = 1, \ldots, X_0 = 1) \cdots \mathbb{P}(X_1 = 1 \mid X_0 = 1)
\end{aligned}$$

$$
\begin{aligned}
&= (\alpha + \gamma) \cdot \beta^{k-1} \\
&= (1-\beta)\beta^{k-1},
\end{aligned}
$$

so T has a geometric distribution with parameter $1-\beta$.

In general, consider a finite-state Markov chain with states $0, \ldots, M-1, M, \ldots, N$, where $0, \ldots, M-1$ are transient and $M, \ldots, N$ are absorbing. By the canonical decomposition,

$$\mathbf{P} = \begin{pmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0}_{(N-M+1)\times M} & \mathbf{I}_{N-M+1} \end{pmatrix}.$$

Let

$$T = \min\{n \geq 0 \mid X_n \in \{M, \ldots, N\}\}$$

be the time until absorption. For $i = 0, \ldots, M-1$ and $j = M, \ldots, N$, the absorption probability is

$$
\begin{aligned}
U_{ij} &:= \mathbb{P}(X_T = j \mid X_0 = i) \\
&= \sum_{k=0}^{M-1} \mathbb{P}(X_T = j \mid X_1 = k) \cdot \mathbb{P}(X_1 = k \mid X_0 = i) \\
&= \sum_{k=0}^{M-1} \mathbb{P}(X_T = j \mid X_0 = k) \cdot Q_{ik} + \mathbb{P}(X_T = j \mid X_1 = j) \cdot \mathbb{P}_{ij} \\
&= \sum_{k=0}^{M-1} U_{kj} Q_{ik} + R_{ij}.
\end{aligned}
$$

In matrix notation, let $\mathbf{U} = [U_{ij}]$, an $M \times (N-M+1)$ matrix. Then

$$\mathbf{U} = \mathbf{R} + \mathbf{QU} \qquad \text{and} \qquad \mathbf{U} = (\mathbf{I} - \mathbf{Q})^{-1}\mathbf{R}.$$

To prove that $\mathbf{I} - \mathbf{Q}$ is invertible, let $\boldsymbol{x}$ satisfy

$$(\mathbf{I} - \mathbf{Q})\boldsymbol{x} = \mathbf{0}_M.$$

Then $\mathbf{Q}\boldsymbol{x} = \boldsymbol{x}$, so for every $n \geq 1$,

$$\mathbf{Q}^n \boldsymbol{x} = \boldsymbol{x}.$$

Since $\{0, \ldots, M-1\}$ is the transient block, we have $\mathbf{Q}^n \to \mathbf{0}_{M\times M}$ as $n \to \infty$. Therefore

$$\boldsymbol{x} = \lim_{n\to\infty} \mathbf{Q}^n \boldsymbol{x} = \mathbf{0}_M.$$

Thus $\ker(\mathbf{I}-\mathbf{Q}) = \{\mathbf{0}_M\}$, so $\mathbf{I}-\mathbf{Q}$ is nonsingular, and hence invertible. For $i = 0, \dots, M-1$, the expected time to absorption is

$$V_i = \mathbb{E}\left[T \mid X_0 = i\right] = \sum_{j=0}^{M-1} \mathbb{P}_{ij}\big(1 + \mathbb{E}\left[T \mid X_0 = j\right]\big) + \sum_{j=M}^{N} 1 \cdot \mathbb{P}_{ij} = 1 + \sum_{j=0}^{M-1} \mathbb{P}_{ij} V_j.$$

In matrix form, if $\boldsymbol{v} = (V_0, \dots, V_{M-1})^\top$, then

$$\boldsymbol{v} = \mathbf{1}_M + \mathbf{Q}\boldsymbol{v} \qquad \text{and} \qquad \boldsymbol{v} = (\mathbf{I}-\mathbf{Q})^{-1}\mathbf{1}_M.$$

We are also interested in the mean number of visits to a transient state k before absorption:

$$\begin{aligned} S_{ik} &:= \mathbb{E}\left[\sum_{n=0}^{T-1} \mathbb{1}_{\{X_n = k\}} \mid X_0 = i\right] \\ &= \delta_{ik} + \sum_{j=0}^{M-1} \mathbb{E}\left[\sum_{n=1}^{T-1} \mathbb{1}_{\{X_n = k\}} \mid X_1 = j\right] \cdot \mathbb{P}_{ij} \\ &= \delta_{ik} + \sum_{j=0}^{M-1} S_{jk} \cdot \mathbb{P}_{ij}. \end{aligned}$$

In matrix notation,

$$\mathbf{S} = \mathbf{I} + \mathbf{Q}\mathbf{S} \qquad \text{and} \qquad \mathbf{S} = (\mathbf{I}-\mathbf{Q})^{-1}.$$

Notice that

$$\sum_{k=0}^{M-1} S_{ik} = \mathbb{E}\left[\sum_{k=0}^{M-1}\sum_{n=0}^{T-1} \mathbb{1}_{\{X_n = k\}} \mid X_0 = i\right] = \mathbb{E}\left[\sum_{n=0}^{T-1}\sum_{k=0}^{M-1} \mathbb{1}_{\{X_n = k\}} \mid X_0 = i\right] = \mathbb{E}\left[T \mid X_0 = i\right] = V_i.$$

This explains why $\boldsymbol{v} = (\mathbf{I}-\mathbf{Q})^{-1}\mathbf{1}_M = \mathbf{S}\mathbf{1}_M$. For transient states $0 \le i, j \le M-1$, let

$$H_j = \inf\{n \ge 0 : X_n = j\} \qquad \text{and} \qquad W_{ij} = \mathbb{P}(H_j < T \mid X_0 = i).$$

Thus $W_{jj} = 1$; when $i \ne j$, W_{ij} is the probability of visiting j at some time $1 \le n \le T-1$. By the restart principle described above, conditional on $H_j < T$, the expected number of visits to j from that time onward is S_{jj}. Conditional on $H_j \ge T$, there are no visits to j. Therefore

$$S_{ij} = W_{ij} S_{jj},$$

and hence

$$W_{ij} = \frac{S_{ij}}{S_{jj}}.$$

We now derive the exact distribution of

$$T = \min\{n \geq 0 \mid X_n \in S_A\}.$$

Assume that we start with a general initial distribution

$$\widetilde{\boldsymbol{\alpha}}_0 = (\alpha_{0,0}, \alpha_{0,1}, \ldots, \alpha_{0,M-1}, \alpha_{0,M}, \ldots, \alpha_{0,N})^\top,$$

and write $\boldsymbol{\alpha}_T = (\alpha_{0,0}, \ldots, \alpha_{0,M-1})^\top$ for its transient-state subvector. Let

$$S_T := \{0, \ldots, M-1\} \qquad \text{and} \qquad S_A := \{M, \ldots, N\}.$$

Then, for $k \geq 1$,

$$\begin{aligned}
\mathbb{P}(T = k) &= \sum_{x_0,\ldots,x_{k-1} \in S_T} \sum_{x_k \in S_A} \mathbb{P}(X_0 = x_0, \ldots, X_k = x_k) \\
&= \sum_{x_0,\ldots,x_{k-1} \in S_T} \sum_{x_k \in S_A} \alpha_{0,x_0} \left(\prod_{i=1}^{k-1} \mathbb{P}_{x_{i-1},x_i} \right) \mathbb{P}_{x_{k-1},x_k} \\
&= \sum_{x_0,\ldots,x_{k-1} \in S_T} \alpha_{0,x_0} \left(\prod_{i=1}^{k-1} \mathbb{P}_{x_{i-1},x_i} \right) \left(\sum_{x_k \in S_A} \mathbb{P}_{x_{k-1},x_k} \right) \\
&= \boldsymbol{\alpha}_T^\top \mathbf{Q}^{k-1} \boldsymbol{q}.
\end{aligned}$$

Here

$$\boldsymbol{q} = \begin{pmatrix} \sum_{j=M}^N \mathbb{P}_{0,j} \\ \vdots \\ \sum_{j=M}^N \mathbb{P}_{M-1,j} \end{pmatrix} = \begin{pmatrix} \mathbb{P}(X_1 \in S_A \mid X_0 = 0) \\ \vdots \\ \mathbb{P}(X_1 \in S_A \mid X_0 = M-1) \end{pmatrix}.$$

Lecture 11 — Wednesday, October 15, 2014

Discrete Phase-Type Distributions

For an absorbing chain with

$$\mathbf{P} = \begin{pmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0}_{(N-M+1)\times M} & \mathbf{I}_{N-M+1} \end{pmatrix}$$

and

$$T = \min\{n \geq 0 \mid X_n \in \{M, \dots, N\}\},$$

the absorption probability is

$$U_{ij} = \mathbb{P}(X_T = j \mid X_0 = i),$$

with $\mathbf{U} = (\mathbf{I}-\mathbf{Q})^{-1}\mathbf{R}$. The expected absorption time $V_i = \mathbb{E}[T \mid X_0 = i]$ satisfies $\boldsymbol{v} = (\mathbf{I}-\mathbf{Q})^{-1}\boldsymbol{e}_M$. The mean number of visits to a transient state is

$$S_{ik} = \mathbb{E}\left[\sum_{n=0}^{T-1} \mathbb{1}_{\{X_n = k\}} \mid X_0 = i\right],$$

and $\mathbf{S} = (\mathbf{I}-\mathbf{Q})^{-1}$. The exact distribution of T is

$$\mathbb{P}(T = k) = \boldsymbol{\alpha}_0^\top \mathbf{Q}^{k-1}\boldsymbol{q},$$

where

$$\boldsymbol{q} = \left(\sum_{j=M}^{N} \mathbb{P}_{0,j}, \dots, \sum_{j=M}^{N} \mathbb{P}_{M-1,j}\right)^T.$$

For example,

$$\mathbb{P}(T = 0) = \alpha_{0,M} + \cdots + \alpha_{0,N} = 1 - \sum_{i=0}^{M-1} \alpha_{0,i}.$$

Definition 11.1 Such a distribution is called a **discrete phase-type distribution**, denoted by $T \sim DPH(\boldsymbol{\alpha}_0, \mathbf{Q})$, where $M = \dim(\mathbf{Q})$ is the number of transient states, $\boldsymbol{\alpha}_0$ is the transient part of the initial distribution and satisfies $\boldsymbol{\alpha}_0 \geq \mathbf{0}$ and $\boldsymbol{\alpha}_0^\top \boldsymbol{e}_M \leq 1$, and $\mathbf{Q}$ is a substochastic transient-to-transient matrix with spectral radius less than 1. The missing initial mass $1 - \boldsymbol{\alpha}_0^\top \boldsymbol{e}_M$ is the probability $\mathbb{P}(T = 0)$.

Notice that $\alpha_{0,M} + \cdots + \alpha_{0,N} = 1 - (\alpha_{0,0} + \cdots + \alpha_{0,M-1})$ is determined by $\boldsymbol{\alpha}_0$. Similarly

$$q_i = \sum_{j=M}^{N} \mathbb{P}_{ij} = 1 - \sum_{j=0}^{M-1} \mathbb{P}_{ij} = 1 - \sum_{j=0}^{M-1} Q_{ij},$$

so $\boldsymbol{q}$ is determined by $\mathbf{Q}$; in particular, $\boldsymbol{q} = (\mathbf{I}-\mathbf{Q})\boldsymbol{e}_M$. Thus

$$\mathbf{P} = \begin{pmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0}_{(N-M+1)\times M} & \mathbf{I}_{N-M+1} \end{pmatrix}$$

can be represented, after combining all absorbing states into one, as

$$\begin{pmatrix} \mathbf{Q} & (\mathbf{I}-\mathbf{Q})\boldsymbol{e}_M \\ \mathbf{0}_{1\times M} & 1 \end{pmatrix}.$$

Revisiting Example 10.1,

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 \\ \alpha & \beta & \gamma \\ 0 & 0 & 1 \end{pmatrix}.$$

After rearranging states,

$$\mathbf{P} = \begin{pmatrix} \beta & \alpha & \gamma \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

so $\mathbf{Q} = [\beta]$ and $\mathbf{R} = [\alpha\ \gamma]$.

Combining the two absorbing states into one gives

$$\mathbf{P} = \begin{pmatrix} \beta & \alpha+\gamma \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \beta & 1-\beta \\ 0 & 1 \end{pmatrix}.$$

Let $\boldsymbol{\alpha}_0 = p$, so $\alpha_{0,0} + \alpha_{0,1} = 1 - p$.

Here $\mathbf{Q} = \beta$, so $\boldsymbol{q} = 1 - \beta$, and therefore

$$\mathbb{P}(T = k) = p\beta^{k-1}(1-\beta).$$

Also $\mathbb{P}(T = 0) = 1 - p$. We call the distribution of T a **zero-modified geometric distribution**. In particular, if $p = \beta$, we obtain the ordinary geometric distribution. Thus, for example, $\mathrm{Geo}(1-p) = DPH_1(p, p)$.

Now let $T \sim DPH_M(\boldsymbol{\alpha}_0, \mathbf{Q})$. For $k = 1, 2, \ldots$, its cumulative distribution function is

$$\begin{aligned} \mathbb{P}(T \le k) &= 1 - \mathbb{P}(T > k) \\ &= 1 - \sum_{j=k+1}^{\infty} \mathbb{P}(T = j) \\ &= 1 - \sum_{j=k+1}^{\infty} \boldsymbol{\alpha}_0^\top \mathbf{Q}^{j-1} \boldsymbol{q} \\ &= 1 - \boldsymbol{\alpha}_0^\top \mathbf{Q}^k (\mathbf{I} - \mathbf{Q})^{-1} \boldsymbol{q} \end{aligned}$$

$$= 1 - \boldsymbol{\alpha}_0^\top \mathbf{Q}^k \boldsymbol{e}_M.$$

The tail probability is $\mathbb{P}(T > k) = \boldsymbol{\alpha}_0^\top \mathbf{Q}^k \boldsymbol{e}_M$. Since T is a nonnegative integer-valued random variable,

$$\mathbb{E}[T] = \sum_{k=0}^{\infty} \mathbb{P}(T > k) = \sum_{k=0}^{\infty} \boldsymbol{\alpha}_0^\top \mathbf{Q}^k \boldsymbol{e}_M = \boldsymbol{\alpha}_0^\top \left(\sum_{k=0}^{\infty} \mathbf{Q}^k \right) \boldsymbol{e}_M = \boldsymbol{\alpha}_0^\top (\mathbf{I} - \mathbf{Q})^{-1} \boldsymbol{e}_M = \boldsymbol{\alpha}_0^\top \boldsymbol{v}$$

where

$$\boldsymbol{v} = (\mathbb{E}[T|X_0 = 0], \ldots, \mathbb{E}[T|X_0 = M-1])^T.$$

This agrees with the result obtained earlier using a first-step analysis.

The class of discrete phase-type distributions is closed under convolution: the sum of two independent discrete phase-type random variables is again discrete phase-type.

Let $X \sim DPH_M(\boldsymbol{\alpha}_0, \mathbf{S})$ and $Y \sim DPH_N(\boldsymbol{\beta}_0, \mathbf{T})$, with X and Y independent. Define $\boldsymbol{s}' = (\mathbf{I} - \mathbf{S})\boldsymbol{e}_M$ and $\boldsymbol{t}' = (\mathbf{I} - \mathbf{T})\boldsymbol{e}_N$, and let $Z = X + Y$. To see that Z is also discrete phase-type, consider a Markov chain with transition probability matrix

$$\mathbf{P} = \begin{pmatrix} \mathbf{S} & \boldsymbol{s}'\boldsymbol{\beta}_0^\top & \boldsymbol{s}'(1 - \boldsymbol{\beta}_0^\top \boldsymbol{e}_N) \\ \mathbf{0}_{N\times M} & \mathbf{T} & \boldsymbol{t}' \\ \mathbf{0}_{1\times M} & \mathbf{0}_{1\times N} & 1 \end{pmatrix}$$

This ensures that once X is absorbed, the initial phase distribution for Y is $\boldsymbol{\beta}_0$. The transient block is

$$\mathbf{Q}_Z = \begin{pmatrix} \mathbf{S} & \boldsymbol{s}'\boldsymbol{\beta}_0^\top \\ \mathbf{0}_{N\times M} & \mathbf{T} \end{pmatrix}.$$

Let $a_0 := 1 - \boldsymbol{\alpha}_0^\top \boldsymbol{e}_M = \mathbb{P}(X = 0)$. The transient initial distribution for Z is

$$\boldsymbol{\delta} = \begin{pmatrix} \boldsymbol{\alpha}_0 \\ a_0 \boldsymbol{\beta}_0 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha}_0 \\ (1 - \boldsymbol{\alpha}_0^\top \boldsymbol{e}_M)\boldsymbol{\beta}_0 \end{pmatrix},$$

where the second block accounts for the case $X = 0$ at time 0, so Z starts directly in the transient phases of Y. (If $\boldsymbol{\alpha}_0^\top \boldsymbol{e}_M = 1$, then this second block is $\mathbf{0}_N$.) The result is

$$Z \sim DPH_{M+N}(\boldsymbol{\delta}, \mathbf{Q}_Z).$$

By induction, the sum of a finite number of independent and identically distributed discrete phase-type random variables is again discrete phase-type.

The mixture of two discrete phase-type distributions is also discrete phase-type. Let $X \sim DPH_M(\boldsymbol{\alpha}_0, \mathbf{S})$ and $Y \sim DPH_N(\boldsymbol{\beta}_0, \mathbf{T})$. If

$$Z = \begin{cases} X, & \text{with probability } p, \\ Y, & \text{with probability } 1-p, \end{cases}$$

and the choice between X and Y is independent of their values, then Z is discrete phase-type. To see this, define

$$\widetilde{\mathbf{Q}} = \begin{pmatrix} \mathbf{S} & \mathbf{0}_{M\times N} \\ \mathbf{0}_{N\times M} & \mathbf{T} \end{pmatrix} \qquad \text{and} \qquad \widetilde{\boldsymbol{\alpha}}_0 = \big(p\boldsymbol{\alpha}_0, (1-p)\boldsymbol{\beta}_0\big),$$

and

$$\widetilde{\boldsymbol{q}} = \begin{pmatrix} (\mathbf{I}-\mathbf{S})\boldsymbol{e}_M \\ (\mathbf{I}-\mathbf{T})\boldsymbol{e}_N \end{pmatrix},$$

where $\boldsymbol{e}_M$ and $\boldsymbol{e}_N$ are the all-ones column vectors of dimensions M and N, respectively.

For $k \geq 1$,

$$\begin{aligned} \mathbb{P}(Z=k) &= p\,\mathbb{P}(X=k) + (1-p)\,\mathbb{P}(Y=k) \\ &= p\,\boldsymbol{\alpha}_0^\top \mathbf{S}^{k-1}(\mathbf{I}-\mathbf{S})\boldsymbol{e}_M + (1-p)\,\boldsymbol{\beta}_0^\top \mathbf{T}^{k-1}(\mathbf{I}-\mathbf{T})\boldsymbol{e}_N \\ &= \widetilde{\boldsymbol{\alpha}}_0^\top\, \widetilde{\mathbf{Q}}^{k-1}\, \widetilde{\boldsymbol{q}}. \end{aligned}$$

Also,

$$\begin{aligned} \mathbb{P}(Z=0) &= p\,\mathbb{P}(X=0) + (1-p)\,\mathbb{P}(Y=0) \\ &= p\,(1-\boldsymbol{\alpha}_0^\top \boldsymbol{e}_M) + (1-p)\,(1-\boldsymbol{\beta}_0^\top \boldsymbol{e}_N) \\ &= 1-\widetilde{\boldsymbol{\alpha}}_0^\top \widetilde{\boldsymbol{e}}_{M+N}, \end{aligned}$$

where $\widetilde{\boldsymbol{e}}_{M+N}$ is the all-ones vector in dimension $M+N$. Therefore $Z \sim DPH_{M+N}(\widetilde{\boldsymbol{\alpha}}_0, \widetilde{\mathbf{Q}})$.

Lecture 12 — Monday, October 20, 2014

A discrete phase-type distribution is the absorption time distribution for a finite-state Markov chain. If $T \sim DPH_m(\boldsymbol{\alpha}_0, \mathbf{Q})$, where $\boldsymbol{\alpha}_0$ is the initial distribution on the transient states, m is the number of transient states, and $\mathbf{Q}$ is the transient-to-transient transition probability matrix, then

$$\mathbb{P}(T = k) = \boldsymbol{\alpha}_0^\top \mathbf{Q}^{k-1} \boldsymbol{q} \qquad \text{and} \qquad \mathbb{P}(T = 0) = 1 - \boldsymbol{\alpha}_0^\top \boldsymbol{e}_m,$$

where $\boldsymbol{q} = (\mathbf{I} - \mathbf{Q})\boldsymbol{e}_m$. Its cumulative distribution function is

$$\mathbb{P}(T \le k) = 1 - \boldsymbol{\alpha}_0^\top \mathbf{Q}^k \boldsymbol{e}_m,$$

and

$$\mathbb{E}[T] = \boldsymbol{\alpha}_0^\top (\mathbf{I} - \mathbf{Q})^{-1} \boldsymbol{e}_m = \boldsymbol{\alpha}_0^\top \boldsymbol{v} \qquad \text{and} \qquad \boldsymbol{v} = (\mathbf{I} - \mathbf{Q})^{-1} \boldsymbol{e}_m.$$

We saw that discrete phase-type distributions are stable under independent sums and finite mixtures. Moreover, any discrete distribution on the nonnegative integers with finite support is discrete phase-type. As a result, we can approximate any discrete distribution on $\mathbb{Z}^+$ arbitrarily closely. If $\boldsymbol{p} = (p_0, p_1, \dots)$ satisfies

$$\sum_{i=0}^{\infty} p_i = 1,$$

then set $p^* = \sum_{k>n} p_k$ and place this remaining mass at $n + 1$. The resulting finite-support distribution $(p_0, p_1, \dots, p_n, p^*)$ can be made arbitrarily close to $\boldsymbol{p}$ by taking n large. In this sense, discrete phase-type distributions are dense in the set of all discrete distributions on $\mathbb{Z}^+$. Discrete phase-type representations are also not unique: the same random variable may admit representations $DPH_m(\boldsymbol{\alpha}_0, \mathbf{Q})$ and $DPH_m(\boldsymbol{\beta}_0, \mathbf{Q}')$.

The absorption results can also be used to solve other problems that do not initially look like absorption problems:

1. To find the expected time until the first visit to a state i, only the behaviour before reaching i matters. Thus the transition probabilities out of i do not matter, and we can make state i absorbing by setting $\mathbb{P}_{ii} = 1$ and $\mathbb{P}_{i\ell} = 0$ for $\ell \neq i$.
2. To find the expected number of visits to state j before visiting state i, make state i absorbing by changing the ith row of $\mathbf{P}$ to be $(0, \dots, 0, 1, 0, \dots)$.

This reduces the problem to finding the expected number of visits to state j before absorption, or the probability of visiting state i before visiting state j.

2. Poisson Processes

Definition and Basic Properties

Definition 12.1 A counting process $\{N(t), t \geq 0\}$ is a **Poisson process** with rate $\lambda > 0$ if the following hold:

1. $N(0) = 0$.
2. $N(t)$ has **independent increments**: for every finite collection of pairwise disjoint time intervals, the corresponding event counts are mutually independent. In particular, $N(t_2) - N(t_1)$ and $N(t_4) - N(t_3)$ are independent for $0 \leq t_1 < t_2 \leq t_3 < t_4$.
3. For all $0 \leq s < t$, the random variable $N(t) - N(s)$ follows a Poisson distribution with parameter $\lambda(t - s)$. That is,

$$\mathbb{P}(N(t) - N(s) = n) = e^{-\lambda(t-s)} \frac{(\lambda(t-s))^n}{n!}, \quad n = 0, 1, 2, \ldots.$$

We can regard the Poisson process $\{N(t), t \geq 0\}$ as the number of random points that fall into the interval $(0, t]$. Under this interpretation, Condition (1) holds naturally, and Condition (2) is equivalent to saying that the numbers of points falling into disjoint intervals are independent.

Proposition 12.2 For a counting process satisfying $N(0) = 0$ and independent increments, Condition (3) holds if and only if the process has stationary increments and, as $h \downarrow 0$,

$$\mathbb{P}(N(h) = 1) = \lambda h + o(h),$$

where

$$\lim_{h \downarrow 0} \frac{o(h)}{h} = 0$$

and $\mathbb{P}(N(h) \geq 2) = o(h)$.

Proof. *Forward implication.* First suppose that $N(t) - N(s) \sim \text{Poisson}(\lambda(t - s))$. Since $\lambda(t - s)$ is only a function of $t - s$, we have stationary increments. Moreover, as $h \downarrow 0$,

$$\mathbb{P}(N(h) - N(0) = 1) = e^{-\lambda h}(\lambda h) = \left(1 - \lambda h + \frac{(\lambda h)^2}{2} - \cdots\right)(\lambda h) = (1 + o(1))\lambda h = \lambda h + o(h)$$

and

$$\begin{aligned}
\mathbb{P}(N(h) - N(0) \geq 2) &= 1 - \mathbb{P}(N(h) - N(0) = 1) - \mathbb{P}(N(h) - N(0) = 0) \\
&= 1 - e^{-\lambda h}\lambda h - e^{-\lambda h} \\
&= e^{-\lambda h}\left(e^{\lambda h} - \lambda h - 1\right) \\
&= e^{-\lambda h}\left(1 + \lambda h + \frac{(\lambda h)^2}{2} + \frac{(\lambda h)^3}{6} + \cdots - \lambda h - 1\right) \\
&= e^{-\lambda h}(o(h)) \\
&= o(h).
\end{aligned}$$

Reverse implication. Define $\mathbb{P}_n(t) = \mathbb{P}(N(t) = n)$. For $\mathbb{P}_0(t)$, we have:

$$\begin{aligned}
\mathbb{P}_0(t+h) &= \mathbb{P}(N(t+h) = 0) \\
&= \mathbb{P}(N(t) = 0, N(t+h) - N(t) = 0) \\
&= \mathbb{P}(N(t) = 0) \cdot \mathbb{P}(N(t+h) - N(t) = 0) \quad \text{(by independent increments)} \\
&= \mathbb{P}_0(t) \cdot (1 - (\lambda h + o(h))) \\
&= \mathbb{P}_0(t) - \lambda h \mathbb{P}_0(t) + o(h).
\end{aligned}$$

Thus

$$\mathbb{P}_0'(t) = \lim_{h \downarrow 0} \frac{\mathbb{P}_0(t+h) - \mathbb{P}_0(t)}{h} = -\lambda \mathbb{P}_0(t).$$

This is a first-order linear ordinary differential equation. Solving for $\mathbb{P}_0(t)$ gives the general solution $ce^{-\lambda t}$. Using the boundary condition

$$\mathbb{P}_0(0) = \mathbb{P}(N(0) = 0) = 1,$$

we get $\mathbb{P}_0(t) = \mathbb{P}(N(t) = 0) = e^{-\lambda t}$.

In general, for $n \geq 1$:

$$\begin{aligned}
\mathbb{P}_n(t+h) &= \mathbb{P}(N(t+h) = n) \\
&= \mathbb{P}(N(t) = n, N(t+h) - N(t) = 0) \\
&\quad + \mathbb{P}(N(t) = n-1, N(t+h) - N(t) = 1) \\
&\quad + \mathbb{P}(N(t+h) = n, N(t+h) - N(t) \geq 2) \\
&= \mathbb{P}_n(t) \cdot \mathbb{P}_0(h) + \mathbb{P}_{n-1}(t) \cdot \mathbb{P}_1(h) + o(h) \\
&= \mathbb{P}_n(t) \cdot (1 - \lambda h) + \mathbb{P}_{n-1}(t) \cdot \lambda h + o(h).
\end{aligned}$$

Therefore

$$\mathbb{P}_n'(t) = -\lambda \mathbb{P}_n(t) + \lambda \mathbb{P}_{n-1}(t),$$

or equivalently

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(e^{\lambda t}\mathbb{P}_n(t)\right) = \lambda e^{\lambda t}\mathbb{P}_{n-1}(t).$$

Integrating from 0 to t gives

$$e^{\lambda t}\mathbb{P}_n(t) - \mathbb{P}_n(0) = \lambda \int_0^t e^{\lambda u}\mathbb{P}_{n-1}(u)\,\mathrm{d}u.$$

For $n \geq 1$, $\mathbb{P}_n(0) = \mathbb{P}(N(0) = n) = 0$, so

$$e^{\lambda t}\mathbb{P}_n(t) = \lambda \int_0^t e^{\lambda u}\mathbb{P}_{n-1}(u)\,\mathrm{d}u.$$

For $n = 1$,

$$e^{\lambda t}\mathbb{P}_1(t) = \lambda \int_0^t e^{\lambda u}\mathbb{P}_0(u)\,\mathrm{d}u = \lambda \int_0^t 1\,\mathrm{d}u = \lambda t,$$

so $\mathbb{P}_1(t) = e^{-\lambda t}(\lambda t)$. For an induction step, assume for some $n \geq 2$ that

$$\mathbb{P}_{n-1}(t) = e^{-\lambda t}\frac{(\lambda t)^{n-1}}{(n-1)!}.$$

Then

$$e^{\lambda t}\mathbb{P}_n(t) = \lambda \int_0^t e^{\lambda u}\mathbb{P}_{n-1}(u)\,\mathrm{d}u = \lambda \int_0^t \frac{(\lambda u)^{n-1}}{(n-1)!}\,\mathrm{d}u = \frac{(\lambda t)^n}{n!}.$$

Hence

$$\mathbb{P}_n(t) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}, \quad n = 0, 1, 2, \ldots.$$

Therefore $N(t) \sim \text{Poisson}(\lambda t)$ for every $t \geq 0$. By stationary increments,

$$N(t) - N(s) \stackrel{d}{=} N(t-s), \quad 0 \leq s < t,$$

so

$$N(t) - N(s) \sim \text{Poisson}(\lambda(t-s)).$$

This is exactly Condition (3). $\square$

Lecture 13 — Wednesday, October 22, 2014

Our earlier absorption results can be used to find the expected time until a first visit to state i and the expected number of visits to state j before visiting state i. We now characterize a Poisson process using the interarrival times $X_1, X_2, \ldots,$ namely the distances between consecutive points. These interarrival times are independent and identically distributed exponential random variables.

First,

$$\mathbb{P}(X_1 > t) = \mathbb{P}(N(t) = 0) = e^{-\lambda t} \implies \mathbb{P}(X_1 \le t) = 1 - e^{-\lambda t},$$

so $X_1 \sim \text{Exp}(\lambda)$. The restart principle says that after the first arrival, the shifted process starts afresh: the next waiting time X_2 is independent of X_1 and has distribution $\text{Exp}(\lambda)$. Repeating the argument after successive arrival times shows that the interarrival times are independent and identically distributed $\text{Exp}(\lambda)$ random variables. A complete proof that independent increments at fixed times imply this restart rule at random arrival times requires the strong Markov property, which is beyond the scope of this course. Thus the Poisson process is the counting process of events with independent exponential interarrival times, and $N(t)$ is the total number of events occurring between time 0 and time t.

This dual description is illustrated in Figure 6: the interarrival times determine the event locations, while $N(t)$ counts how many locations have been passed by time t.

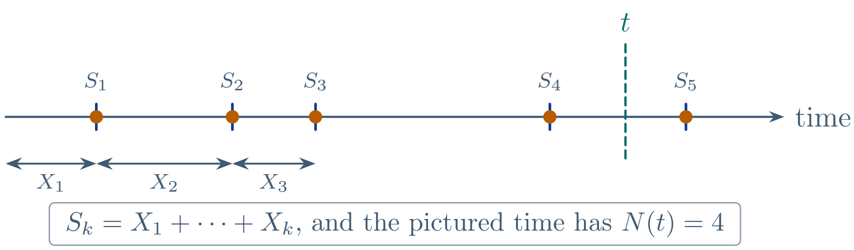


Figure 6

Combining and Thinning Poisson Processes

Theorem 13.1 Let $\{N_1(t) : t \ge 0\}$ and $\{N_2(t) : t \ge 0\}$ be independent Poisson processes with intensities λ_1 and λ_2, respectively. Then $N(t) = N_1(t) + N_2(t)$ is also a Poisson process with intensity $\lambda_1 + \lambda_2$.

Proof. We check the defining properties. First, $N(0) = N_1(0) + N_2(0) = 0$, so Condition (1) is satisfied.

For

$$0 \le t_1 < t_2 \le t_3 < t_4,$$

the increment $N(t_2) - N(t_1)$ is the sum of the corresponding increments of N_1 and N_2, and $N(t_4) - N(t_3)$ is the sum of the corresponding increments over the disjoint interval $(t_3, t_4]$. Since N_1 and N_2 are independent Poisson processes with independent increments, these two increments of N are independent, which is Condition (2).

Finally, the sum of independent Poisson random variables with parameters $\lambda_1(t-s)$ and $\lambda_2(t-s)$ is Poisson with parameter $(\lambda_1 + \lambda_2)(t-s)$, which is Condition (3). Hence N is a Poisson process with intensity $\lambda_1 + \lambda_2$. □

Theorem 13.2 (Combining) If X and Y are independent Poisson random variables with parameters a and b, respectively, then $X + Y$ is Poisson with parameter $a + b$. In general, if $N_1(t), N_2(t), \ldots, N_k(t)$ are independent Poisson processes with intensities $\lambda_1, \ldots, \lambda_k$, then

$$N(t) = \sum_{i=1}^{k} N_i(t)$$

is Poisson with intensity

$$\lambda = \sum_{i=1}^{k} \lambda_i.$$

Proof. Let $X \sim \text{Poisson}(a)$ and $Y \sim \text{Poisson}(b)$ be independent. The moment generating function of a Poisson random variable with parameter θ is

$$M_{\text{Poisson}(\theta)}(t) = \exp\bigl(\theta(e^t - 1)\bigr).$$

Hence

$$M_X(t) = \exp\bigl(a(e^t - 1)\bigr) \qquad \text{and} \qquad M_Y(t) = \exp\bigl(b(e^t - 1)\bigr).$$

By independence,

$$\begin{aligned} M_{X+Y}(t) &= \mathbb{E}\bigl[e^{t(X+Y)}\bigr] = \mathbb{E}\bigl[e^{tX}e^{tY}\bigr] \\ &= \mathbb{E}\bigl[e^{tX}\bigr]\,\mathbb{E}\bigl[e^{tY}\bigr] = M_X(t)M_Y(t) \\ &= \exp\bigl((a+b)(e^t - 1)\bigr). \end{aligned}$$

This is the moment generating function of Poisson$(a+b)$, so $X+Y \sim \text{Poisson}(a+b)$.

Now let $N_1, \ldots, N_k$ be independent Poisson processes with intensities $\lambda_1, \ldots, \lambda_k$, and define

$$N(t) = \sum_{i=1}^{k} N_i(t).$$

For each fixed $t \geq 0$, the random variables $N_i(t)$ are independent with

$$N_i(t) \sim \text{Poisson}(\lambda_i t).$$

Applying the result above repeatedly yields

$$N(t) \sim \text{Poisson}\left(t \sum_{i=1}^{k} \lambda_i\right).$$

Also, $N(0) = 0$, and N has independent stationary increments because each N_i has independent stationary increments and the processes are independent across i. Therefore N is a Poisson process with intensity

$$\lambda = \sum_{i=1}^{k} \lambda_i.$$

□

Theorem 13.3 (Thinning/Splitting) Suppose that each event of a Poisson process $N(t)$ with intensity λ is independently assigned to type 1 with probability p and to type 2 with probability $1-p$. Let $N_1(t)$ and $N_2(t)$ count the type 1 and type 2 events, respectively. Then N_1 and N_2 are independent Poisson processes with intensities $p\lambda$ and $(1-p)\lambda$.

Combining adds intensities, while thinning partitions one intensity according to the assignment probabilities; Figure 7 places the two operations side by side.

Proof. First, $N_1(0) = 0$. The process $N_1(t)$ has independent increments as a direct consequence of the independent increments property of $N(t)$, and it has stationary increments because $N(t)$ has stationary increments and the splitting rule does not change over time. Moreover,

$$\begin{aligned}
\mathbb{P}(N_1(h) = 1) &= \mathbb{P}(N_1(h) = 1 \mid N(h) = 1) \cdot \mathbb{P}(N(h) = 1) \\
&\quad + \mathbb{P}(N_1(h) = 1 \mid N(h) \geq 2) \cdot \mathbb{P}(N(h) \geq 2) \\
&= p(\lambda h + o(h)) + \mathbb{P}(N_1(h) = 1 \mid N(h) \geq 2) \cdot o(h)
\end{aligned}$$

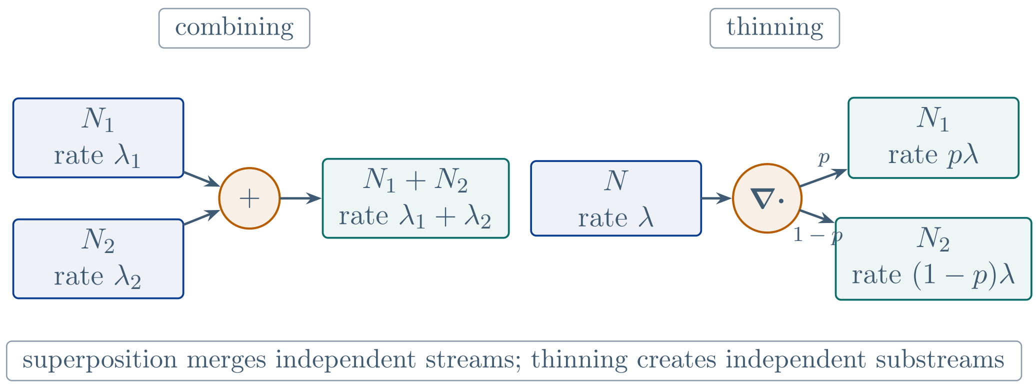


FIGURE 7

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

Also

$$\mathbb{P}(N_1(h) \geq 2) \leq \mathbb{P}(N(h) \geq 2) = o(h).$$

Thus $N_1(t)$ is a Poisson process with intensity $p\lambda$. Similarly, $N_2(t)$ is a Poisson process with intensity $(1-p)\lambda$.

Lecture 14 — Monday, October 27, 2014

It remains to show that N_1 and N_2 are independent. Observe that

$$\begin{aligned}
\mathbb{P}(N_1(t) = m, N_2(t) = n) &= \mathbb{P}(N_1(t) = m, N_2(t) = n \mid N(t) = m+n) \cdot \mathbb{P}(N(t) = m+n) \\
&= \binom{m+n}{m} p^m (1-p)^n \cdot e^{-\lambda t} \frac{(\lambda t)^{m+n}}{(m+n)!} \\
&= \frac{(m+n)!}{m!n!} p^m (1-p)^n e^{-\lambda t} \frac{\lambda^{m+n} t^{m+n}}{(m+n)!} \\
&= \left(\frac{(p\lambda t)^m}{m!} e^{-p\lambda t} \right) \cdot \left(\frac{((1-p)\lambda t)^n}{n!} e^{-(1-p)\lambda t} \right) \\
&= \mathbb{P}(N_1(t) = m) \cdot \mathbb{P}(N_2(t) = n).
\end{aligned}$$

Hence $N_1(t)$ and $N_2(t)$ are independent for each fixed t. To obtain independence of the processes, note that for any disjoint intervals, the vectors of increments of N_1 and N_2 are formed by independently marking the points of the corresponding increments of N. Therefore the finite-dimensional vectors

$$(N_1(t_1), \ldots, N_1(t_k)) \qquad \text{and} \qquad (N_2(t_1), \ldots, N_2(t_k))$$

are independent for every $0 \le t_1 < \cdots < t_k$. Thus N_1 and N_2 are independent Poisson processes. $\square$

Order Statistics and Simulation

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables. The **order statistics** of $X_1, \ldots, X_n$ are $X_{(1)} = \min\{X_1, \ldots, X_n\}$, $X_{(2)} =$ second smallest of $\{X_1, \ldots, X_n\}$, $\ldots$, $X_{(n)} = \max\{X_1, \ldots, X_n\}$.

Theorem 14.1 Let $\{N(t), t \ge 0\}$ be a Poisson process with intensity λ. Conditional on $N(t) = n$, the points of N in $[0, t]$ are distributed as the order statistics from a sample of size n from the uniform distribution $\text{Unif}(0, t)$. That is,

$$(S_1, \ldots, S_n \mid N(t) = n) \overset{d}{=} (U_{(1)}, \ldots, U_{(n)}),$$

where $U_1, \ldots, U_n \overset{\text{iid}}{\sim} \text{Unif}(0, t)$.

Proof. Let

$$a_1 < b_1 < a_2 < b_2 < \cdots < a_n < b_n \le t.$$

$$\begin{aligned}
\mathbb{P}(S_i \in (a_i, b_i], i = 1, \ldots, n \mid N(t) = n) &= \frac{\mathbb{P}(N(a_1) = 0, N(b_1) - N(a_1) = 1, \ldots, N(t) - N(b_n) = 0)}{\mathbb{P}(N(t) = n)} \\
&= \frac{e^{-\lambda t}\lambda^n \prod_{i=1}^n (b_i - a_i)}{e^{-\lambda t}(\lambda t)^n / n!} \\
&= \frac{n!}{t^n} \prod_{i=1}^n (b_i - a_i).
\end{aligned}$$

Dividing both sides by $\prod_{i=1}^n (b_i - a_i)$ and taking the limit as $b_i \to a_i$, we get

$$f_{S_1, \ldots, S_n \mid N(t) = n}(a_1, \ldots, a_n) = \frac{n!}{t^n} \mathbb{1}_{\{0 < a_1 < \cdots < a_n < t\}},$$

which is the density of $(U_{(1)}, \ldots, U_{(n)})$. $\square$

There are two standard ways to simulate the points of a Poisson process on $[0, t]$. We can simulate independent $\text{Exp}(\lambda)$ interarrival times until their sum is greater than t. Alternatively, we can simulate $N \sim \text{Poisson}(\lambda t)$, then simulate N independent $\text{Unif}[0, t]$ points and arrange them in increasing order.

In many cases, it is helpful to define a Poisson process on all of $\mathbb{R}$ rather than on $[0, \infty)$. Then it is no longer natural to count points starting from 0, since 0 is arbitrary. Instead, we view N as a random set function: for a Borel set $A \subset \mathbb{R}$, let $N(A)$ be the number of points of the Poisson process in A. Most commonly, A is an interval, so $N((a, b])$ is the number of points in $(a, b]$. From a measure-theoretic perspective, N is often called a **Poisson random measure**. Counts on any finite collection of pairwise disjoint Borel sets are mutually independent, and $N(A) \sim \text{Poisson}(\lambda|A|)$ whenever A has finite Lebesgue measure $|A|$.

3. Continuous-Time Markov Chains

Construction and the Discrete Skeleton

Definition 14.2 A stochastic process $\{X(t)\}_{t\geq 0}$ is a **continuous-time Markov chain** if it has right-continuous, piecewise-constant paths and:

1. For $t \geq 0$, $X(t)$ takes values in an at most countable set $S \subseteq \{0, 1, 2, \dots\}$.
2. The Markov property holds: for $s, t \geq 0$ and $i, j \in S$,

$$\mathbb{P}(X(t+s) = j \mid X(s) = i, X(u) = x(u), 0 \leq u < s) = \mathbb{P}(X(t+s) = j \mid X(s) = i)$$

   whenever the conditioning event has positive probability.

The probabilistic properties of the future depend only on the current state, not the past. The chain is called **time-homogeneous** if $\mathbb{P}(X(t+s) = j \mid X(s) = i)$ does not depend on s, which is our default setting. Unless stated otherwise, we also assume that the chain is nonexplosive, so it makes only finitely many jumps on every bounded time interval.

Lecture 15 — Wednesday, October 29, 2014

The time spent in a given state is called the **sojourn time**. To construct a continuous-time Markov chain, we need to answer two questions:

Question 15.1

1. How long does the chain stay in state i?
2. When the chain leaves state i, which state does it enter next?

To answer Question (1), let T_i be the random time that the chain spends in state i. Then

$$\begin{aligned}\mathbb{P}(T_i > t+s \mid T_i > s) &= \mathbb{P}(X(u) = i \text{ for all } s \le u \le s+t \mid X(u) = i \text{ for all } 0 \le u \le s) \\ &= \mathbb{P}(X(u) = i \text{ for all } s \le u \le s+t \mid X(s) = i) \\ &= \mathbb{P}(X(u) = i \text{ for all } 0 \le u \le t \mid X(0) = i) \\ &= \mathbb{P}(T_i > t).\end{aligned}$$

This is the memoryless property, so T_i follows an exponential distribution. Let its parameter be ν_i, so $T_i \sim \mathrm{Exp}(\nu_i)$. Each time the process enters state i, the amount of time it spends there before going to another state is exponentially distributed with mean $1/\nu_i$. If $\nu_i = 0$, then $\mathbb{P}(T_i > t) = 1$ for every t, so $T_i = \infty$ and state i is absorbing.

For Question (2), the standard pure-jump construction gives independence between the sojourn time in state i and the destination of the next jump. We denote the conditional probability that the chain goes to j when it leaves i by $\mathbb{P}_{ij}$.

A continuous-time Markov chain stays in state i for an exponential random time T_i, then jumps to another state j with probability $\mathbb{P}_{ij}$, then stays in state j for an exponential random time T_j, and so on.

The construction separates holding times from the embedded jump chain, as shown in Figure 8. The upper timeline records both pieces of information; the lower chain retains only the successive states.

For every nonabsorbing state i, the jump probabilities satisfy $\mathbb{P}_{ii} = 0$ because $\mathbb{P}_{ij}$ records the next

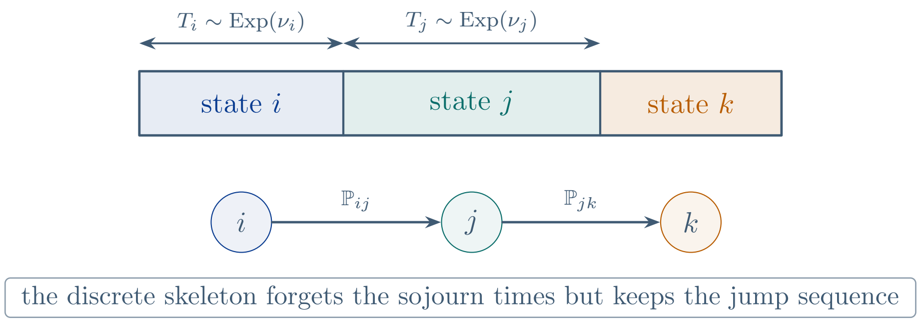


FIGURE 8

state after leaving i, and

$$\sum_{j \neq i} \mathbb{P}_{ij} = 1$$

for every such i. If $\nu_i = 0$, the process never leaves i; to obtain a stochastic matrix for the embedded chain, we use the convention $\mathbb{P}_{ii} = 1$ on that absorbing state. The resulting discrete-time chain governs the successive states visited by the continuous-time chain but does not record sojourn times. It is called the **discrete skeleton** (or the **embedded discrete-time Markov chain**).

Kolmogorov Forward and Backward Equations

We want to understand how

$$\mathbb{P}_{ij}(t) = \mathbb{P}(X(t) = j \mid X(0) = i) = \mathbb{P}(X(t+s) = j \mid X(s) = i)$$

changes as a function of t.

For $i \neq j$, define $q_{ij} = \nu_i \mathbb{P}_{ij}$. Here ν_i is the rate of leaving i, while $\mathbb{P}_{ij}$ is the probability of jumping to j after leaving i. Thus q_{ij} is the "probability flow" from i to j. Notice that

$$\nu_i = \nu_i \sum_{j \neq i} \mathbb{P}_{ij} = \sum_{j \neq i} q_{ij}.$$

Under the standard small-time regularity condition that the probability of two or more jumps from a fixed starting state is $o(h)$, for small $h > 0$,

$$\mathbb{P}_{ii}(h) = \mathbb{P}(X(h) = i \mid X(0) = i) = \mathbb{P}(T_i > h) + o(h)$$

because h is small enough that the probability of two or more transitions is $o(h)$. Hence

$$\mathbb{P}_{ii}(h) = e^{-\nu_i h} + o(h) = (1 - \nu_i h + o(h)) + o(h) = 1 - \nu_i h + o(h).$$

$$\lim_{h \downarrow 0} \frac{\mathbb{P}_{ii}(h) - 1}{h} = -\nu_i. \tag{15.1}$$

For $i \neq j$ and h ever so small,

$$\begin{aligned}
\mathbb{P}_{ij}(h) &= \mathbb{P}(X(h) = j \mid X(0) = i) \\
&= \mathbb{P}_{ij}\mathbb{P}(T_i \leq h) + o(h) \\
&= \mathbb{P}_{ij}(1 - e^{-\nu_i h}) + o(h) \\
&= \mathbb{P}_{ij}(\nu_i h + o(h)) + o(h) \\
&= \mathbb{P}_{ij}\nu_i h + o(h) \\
&= q_{ij} h + o(h),
\end{aligned}$$

so

$$\lim_{h \downarrow 0} \frac{\mathbb{P}_{ij}(h)}{h} = q_{ij}. \tag{15.2}$$

Let $\mathbf{P}(t) = \{\mathbb{P}_{ij}(t)\}_{i,j \in S}$ be the transition probability matrix at time t, with $\mathbf{P}(0) = \mathbf{I}$. Combining (15.1) and (15.2) gives

$$\lim_{h \downarrow 0} \frac{\mathbf{P}(h) - \mathbf{P}(0)}{h} = \begin{pmatrix} -\nu_0 & q_{01} & q_{02} & \cdots \\ q_{10} & -\nu_1 & q_{12} & \cdots \\ q_{20} & q_{21} & -\nu_2 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} =: \mathbf{R}.$$

Definition 15.2 The matrix $\mathbf{R}$ is called the **generator matrix** of the continuous-time Markov chain.

The generator matrix combines the information about sojourn times and transition probabilities. Knowing $\mathbf{R}$ determines $\boldsymbol{\nu}$ and the jump probabilities from every nonabsorbing state, and conversely. If $\mathbf{R} = \{r_{ij}\}_{i,j \in S}$, then

$$r_{ij} = \begin{cases} -\nu_i & i = j \\ q_{ij} & i \neq j \end{cases}.$$

The row sums of $\mathbf{R}$ are always 0.

The Chapman–Kolmogorov equations still hold in continuous time:

$$\begin{aligned}\mathbb{P}_{ij}(t+s) &= \mathbb{P}(X(t+s)=j \mid X(0)=i)\\ &= \sum_{k\in S}\mathbb{P}(X(t+s)=j \mid X(s)=k, X(0)=i)\cdot\mathbb{P}(X(s)=k \mid X(0)=i)\\ &= \sum_{k\in S}\mathbb{P}_{kj}(t)\cdot\mathbb{P}_{ik}(s).\end{aligned}$$

In matrix notation, $\mathbf{P}(t+s) = \mathbf{P}(s)\mathbf{P}(t)$.

Thus

$$\mathbf{P}(t+h) = \mathbf{P}(h)\mathbf{P}(t),$$

so

$$\mathbf{P}(t+h) - \mathbf{P}(t) = (\mathbf{P}(h) - \mathbf{I})\mathbf{P}(t).$$

Dividing by h and taking $h \downarrow 0$ gives

$$\mathbf{P}'(t) = \mathbf{R}\mathbf{P}(t). \tag{15.3}$$

This is the **Kolmogorov backward equation**.

Lecture 16 — Monday, November 03, 2014

Recall that the sojourn time in a nonabsorbing state i has distribution $T_i \sim \mathrm{Exp}(\nu_i)$. Conditional on leaving state i, the probability of going to j is $\mathbb{P}_{ij}$, with $\mathbb{P}_{ii} = 0$. On an absorbing state, we use $\mathbb{P}_{ii} = 1$ for the discrete skeleton. Also, $\mathbf{P}(t) = \{\mathbb{P}_{ij}(t)\}_{i,j\in S}$, and

$$\mathbf{R} = \begin{pmatrix} -\nu_0 & q_{01} & \cdots \\ q_{10} & -\nu_1 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

is the generator matrix, where $q_{ij} = \nu_i \mathbb{P}_{ij}$ and

$$\sum_{j\in S} R_{ij} = 0 \quad \text{for all } i \in S.$$

The Chapman–Kolmogorov equation is $\mathbf{P}(t+s) = \mathbf{P}(t)\mathbf{P}(s)$, and from it we derived the Kolmogorov backward equation $\mathbf{P}'(t) = \mathbf{R}\mathbf{P}(t)$. Similarly,

$$\mathbf{P}(t+h) = \mathbf{P}(t)\mathbf{P}(h) \implies \lim_{h\to 0}\frac{\mathbf{P}(t+h)-\mathbf{P}(t)}{h} = \mathbf{P}(t)\lim_{h\to 0}\frac{\mathbf{P}(h)-\mathbf{I}}{h} \implies \mathbf{P}'(t) = \mathbf{P}(t)\mathbf{R}. \tag{16.1}$$

This is the **Kolmogorov forward equation**.

In entrywise form, the backward equation is

$$P'_{ij}(t) = \sum_{k\in S} R_{ik}\mathbb{P}_{kj}(t) = \sum_{k\neq i} q_{ik}\mathbb{P}_{kj}(t) - \nu_i\mathbb{P}_{ij}(t).$$

and the forward equation is

$$P'_{ij}(t) = \sum_{k\neq j}\mathbb{P}_{ik}(t)q_{kj} - \nu_j\mathbb{P}_{ij}(t).$$

Remark 16.1 In the derivation of the Kolmogorov forward equation, we quietly interchanged the order of a limit and a summation:

$$\lim_{h\downarrow 0}\mathbf{P}(t)\frac{\mathbf{P}(h)-\mathbf{I}}{h} = \mathbf{P}(t)\left(\lim_{h\downarrow 0}\frac{\mathbf{P}(h)-\mathbf{I}}{h}\right).$$

This is automatic when the state space S is finite; for a countably infinite state space it requires additional regularity or summability conditions. The backward equation avoids this particular interchange and is consequently the safer starting point in the general theory.

When the state space is finite, or more generally when the generator defines a bounded operator, the matrix $\mathbf{P}(t)$ solves the differential equation $\mathbf{P}'(t) = \mathbf{R}\mathbf{P}(t)$ with initial condition $\mathbf{P}(0) = \mathbf{I}$. The solution is the **matrix exponential**

$$e^{t\mathbf{R}} := \mathbf{I} + t\mathbf{R} + \frac{t^2\mathbf{R}^2}{2!} + \frac{t^3\mathbf{R}^3}{3!} + \cdots = \sum_{n=0}^{\infty}\frac{t^n}{n!}\mathbf{R}^n.$$

The series converges in any matrix norm in finite dimensions, and in operator norm for a bounded generator. Moreover, $e^{t\mathbf{R}}$ solves a matrix differential equation:

$$\frac{\mathrm{d}}{\mathrm{d}t}e^{t\mathbf{R}} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\mathbf{I} + \sum_{n=1}^{\infty}\frac{t^n}{n!}\mathbf{R}^n\right) = \sum_{n=1}^{\infty}\frac{t^{n-1}}{(n-1)!}\mathbf{R}^n = \mathbf{R}\sum_{n=1}^{\infty}\frac{t^{n-1}}{(n-1)!}\mathbf{R}^{n-1} = \mathbf{R}\sum_{m=0}^{\infty}\frac{t^m}{m!}\mathbf{R}^m = \mathbf{R}e^{t\mathbf{R}}.$$

Also, $e^{t\mathbf{R}}\big|_{t=0} = \mathbf{I}$. Hence $\mathbf{P}(t) = e^{t\mathbf{R}}$ satisfies the backward equation. The forward equation is handled similarly.

Warning The matrix exponential $e^{t\mathbf{R}}$ is *not* obtained by exponentiating each entry of $t\mathbf{R}$ separately!

In general, $e^{t\mathbf{R}}$ is not easy to calculate. However, if $\mathbf{R}$ is diagonalizable, say $\mathbf{R} = \mathbf{B}\mathbf{D}\mathbf{B}^{-1}$ with $\mathbf{B}$ invertible and $\mathbf{D} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, then we have

$$e^{t\mathbf{R}} = \sum_{n=0}^{\infty} \frac{t^n}{n!}\mathbf{R}^n = \sum_{n=0}^{\infty} \frac{t^n}{n!}(\mathbf{B}\mathbf{D}\mathbf{B}^{-1})^n = \sum_{n=0}^{\infty} \frac{t^n}{n!}\mathbf{B}\mathbf{D}^n\mathbf{B}^{-1} = \mathbf{B}\left(\sum_{n=0}^{\infty} \frac{t^n}{n!}\mathbf{D}^n\right)\mathbf{B}^{-1} = \mathbf{B}e^{t\mathbf{D}}\mathbf{B}^{-1}$$

where $e^{t\mathbf{D}}$ *is* computed entrywise along the diagonal:

$$e^{t\mathbf{D}} = \operatorname{diag}(e^{t\lambda_1}, \ldots, e^{t\lambda_n}),$$

which holds because for each $m \geq 0$,

$$\mathbf{D}^m = \operatorname{diag}(\lambda_1^m, \ldots, \lambda_n^m),$$

and therefore

$$\begin{aligned} e^{t\mathbf{D}} &= \sum_{m=0}^{\infty} \frac{t^m}{m!}\mathbf{D}^m \\ &= \sum_{m=0}^{\infty} \frac{t^m}{m!} \operatorname{diag}(\lambda_1^m, \ldots, \lambda_n^m) \\ &= \operatorname{diag}\left(\sum_{m=0}^{\infty} \frac{(t\lambda_1)^m}{m!}, \ldots, \sum_{m=0}^{\infty} \frac{(t\lambda_n)^m}{m!}\right) \\ &= \operatorname{diag}(e^{t\lambda_1}, \ldots, e^{t\lambda_n}). \end{aligned}$$

Birth and Death Processes

A **birth and death process** is a continuous-time stochastic process $X(t)$, which we think of as the number of individuals in a system at time t. At each time, the population can only change by 1 or -1, corresponding to a birth or a death. Thus $\mathbb{P}_{ij} = 0$ if $|i - j| > 1$, equivalently $q_{ij} = 0$ if

$|i-j|>1$. The generator matrix $\mathbf{R}$ has the form

$$\mathbf{R} = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & \dots \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & \dots \\ 0 & \mu_2 & -(\lambda_2+\mu_2) & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

where $\{\lambda_i\}_{i\geq 0}$ are the birth rates and $\{\mu_i\}_{i\geq 1}$ are the death rates. For $i \geq 1$,

$$q_{i,i+1} = \nu_i \mathbb{P}_{i,i+1} = \lambda_i \qquad \text{and} \qquad q_{i,i-1} = \nu_i \mathbb{P}_{i,i-1} = \mu_i.$$

Thus $\nu_i = \lambda_i + \mu_i$. For $i=0$, $\mathbb{P}_{0,1} = 1$. Intuitively, when the system is in state i, the next potential birth occurs after an exponential time with intensity λ_i, and the next potential death occurs after an exponential time with intensity μ_i. These two times "compete" with each other; the smaller time determines the next state change.

The nearest-neighbour structure and the indexing of the two rate families are displayed in Figure 9.

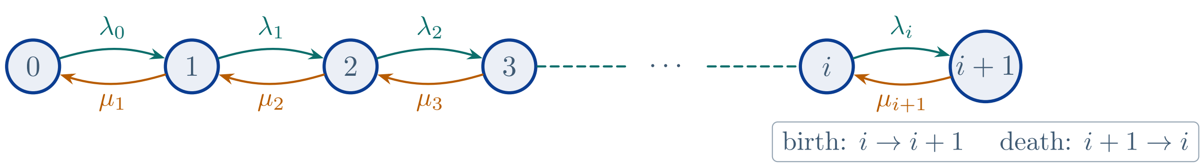


FIGURE 9

If $T_i^{(b)} \sim \text{Exp}(\lambda_i)$ and $T_i^{(d)} \sim \text{Exp}(\mu_i)$ are independent, then

$$\mathbb{P}(T_i^{(b)} < T_i^{(d)}) = \frac{\lambda_i}{\lambda_i+\mu_i} = \mathbb{P}_{i,i+1}$$

and

$$\mathbb{P}(T_i^{(d)} < T_i^{(b)}) = \frac{\mu_i}{\lambda_i+\mu_i} = \mathbb{P}_{i,i-1}$$

Also, $T_i = T_i^{(b)} \wedge T_i^{(d)}$ is the time until the system changes state, and $T_i \sim \text{Exp}(\lambda_i+\mu_i)$ for $i \geq 1$. Hence $\nu_i = \lambda_i + \mu_i$ for $i \geq 1$, while $\nu_0 = \lambda_0$.

Stationary Distributions for Birth and Death Processes

For a finite-state chain, and for the regular nonexplosive countable-state chains considered here when the required sums are finite, a distribution $\boldsymbol{\pi} = (\pi_0, \pi_1, \dots)^\top$ is stationary if $\mathbf{R}^\top \boldsymbol{\pi} = \mathbf{0}_{|S|}$

and

$$\sum_{j \in S} \pi_j = 1.$$

Recall that for a discrete-time Markov chain, $\mathbf{P}^\top \boldsymbol{\pi} = \boldsymbol{\pi}$. Here, the condition $\mathbf{R}^\top \boldsymbol{\pi} = \mathbf{0}_{|S|}$ is equivalent to

$$\sum_{i \in S} \pi_i r_{ij} = 0 \quad \text{for all } j \in S.$$

Assume now that the birth and death process is irreducible, so $\lambda_i > 0$ and $\mu_i > 0$ on their respective ranges. For $j = 0$,

$$-\lambda_0 \pi_0 + \mu_1 \pi_1 = 0 \implies \pi_1 = \frac{\lambda_0}{\mu_1} \pi_0$$

and for $j = 1$,

$$\lambda_0 \pi_0 - (\lambda_1 + \mu_1)\pi_1 + \mu_2 \pi_2 = 0 \implies \pi_2 = \frac{\lambda_1}{\mu_2}\pi_1 = \frac{\lambda_1 \lambda_0}{\mu_2 \mu_1}\pi_0.$$

In general, for $j \geq 1$,

$$\pi_j = \pi_0 \prod_{k=0}^{j-1} \frac{\lambda_k}{\mu_{k+1}}$$

The condition

$$\sum_{j \in S} \pi_j = 1$$

implies

$$\pi_0 \left(1 + \sum_{j=1}^{\infty} \prod_{k=0}^{j-1} \frac{\lambda_k}{\mu_{k+1}} \right) = 1$$

This sum must converge for a stationary distribution to exist; under the standing nonexplosion assumption, convergence also makes the normalized solution stationary.

Lecture 17 — Wednesday, November 05, 2014

Recall the Kolmogorov forward equation $\mathbf{P}'(t) = \mathbf{P}(t)\mathbf{R}$. In entrywise form, the backward equation is

$$P'_{ij}(t) = \sum_{k \neq i} q_{ik} \mathbb{P}_{kj}(t) - \nu_i \mathbb{P}_{ij}(t).$$

The forward equation is

$$P'_{ij}(t) = \sum_{k \neq j} \mathbb{P}_{ik}(t) q_{kj} - \nu_j \mathbb{P}_{ij}(t).$$

Also, $\mathbf{P}(t) = e^{t\mathbf{R}}$. If $\mathbf{R}$ is diagonalizable, $\mathbf{R} = \mathbf{BDB}^{-1}$ for some diagonal $\mathbf{D}$, then $e^{t\mathbf{R}} = \mathbf{B}e^{t\mathbf{D}}\mathbf{B}^{-1}$, where $e^{t\mathbf{D}} = \mathrm{diag}(e^{t\lambda_1}, \ldots, e^{t\lambda_n})$. For a birth and death process,

$$\mathbb{P}_{ij} = 0 \text{ for } |i-j| > 1 \qquad \text{if and only if} \qquad q_{ij} = 0 \text{ for } |i-j| > 1.$$

In this case,

$$\mathbf{R} = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 \\ 0 & \mu_2 & \ddots \end{pmatrix}.$$

Here $\nu_i = \lambda_i + \mu_i$ and $\nu_0 = \lambda_0$, with λ_i the birth rates and μ_i the death rates. Moreover,

$$\mathbb{P}_{i,i+1} = \frac{\lambda_i}{\lambda_i + \mu_i}, \qquad \mathbb{P}_{i,i-1} = \frac{\mu_i}{\lambda_i + \mu_i}, \qquad \text{and} \qquad \mathbb{P}_{0,1} = 1.$$

Because of this structure, it is usually easier to write $\mathbf{R}$ directly than to write ν_i and $\mathbb{P}_{ij}$ first.

Special Cases of Birth and Death Processes

We enumerate several special cases.

1. If $\mu_i = 0$ for $i = 0, 1, 2, \ldots$, this is called a **pure birth process**.
2. More depressingly, if $\lambda_i = 0$ for $i = 0, 1, 2, \ldots$, then there are no births, only deaths. This is called a **pure death process**.
3. If $\mu_i = 0$ and $\lambda_i = \lambda > 0$, then the times between births are independent and exponentially distributed with the same parameter λ, regardless of i. This is the counting process of Poisson events, and hence a Poisson process.
4. If $\mu_i = 0$ and $\lambda_i = i\lambda$, this is a **Yule process**. The birth rate is proportional to the current population. Intuitively, each individual has the same constant birth rate λ, independently of the others. Thus the time until the next birth, given current population i, is the minimum of i independent $\mathrm{Exp}(\lambda)$ random variables, hence has distribution $\mathrm{Exp}(i\lambda)$.
5. If $\lambda_i = i\lambda$ and $\mu_i = i\mu$, this is a **linear growth model**. Individuals are independent and have common birth rate λ and death rate μ.

6. If $\lambda_i = \lambda$ for all $i \geq 0$ and
$$\mu_i = \begin{cases} i\mu, & i < s \\ s\mu, & i \geq s \end{cases},$$
this is an **M/M/s queue**. Each arrival is a birth, and each service completion is a death. For $i < s$, there are i busy servers, so the death rate is the sum of the individual service rates, namely $i\mu$. For $i \geq s$, all s servers are busy, and the death rate is $s\mu$.

7. If $\lambda_i = \lambda$ for all $i \geq 0$ and $\mu_i = \mu$ for all $i \geq 1$, this is the special case of an $M/M/s$ queue with $s = 1$, namely an **M/M/1 queue**.

Consider the forward equations for birth and death processes, $\mathbf{P}'(t) = \mathbf{P}(t)\mathbf{R}$. For the special case of a pure birth process, where $\mu_i = 0$ for $i = 1, 2, \ldots$, we have $\mathbb{P}_{ij}(t) = 0$ if $j < i$. If $j = i$, then $P'_{ii}(t) = -\lambda_i \mathbb{P}_{ii}(t)$, so $\mathbb{P}_{ii}(t) = e^{-\lambda_i t}$. Probabilistically, since the chain is a pure birth process,
$$\mathbb{P}_{ii}(t) = \mathbb{P}(X(t) = i \mid X(0) = i) = \mathbb{P}(T_i > t) = e^{-\lambda_i t}.$$
For $j > i$,
$$\begin{aligned} P'_{ij}(t) &= \lambda_{j-1}\mathbb{P}_{i,j-1}(t) - \lambda_j \mathbb{P}_{ij}(t), \\ P'_{ij}(t) + \lambda_j \mathbb{P}_{ij}(t) &= \lambda_{j-1}\mathbb{P}_{i,j-1}(t). \end{aligned}$$
Multiplying by $e^{\lambda_j t}$ and integrating from 0 to t gives
$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\Big(e^{\lambda_j t}\mathbb{P}_{ij}(t)\Big) &= \lambda_{j-1} e^{\lambda_j t}\mathbb{P}_{i,j-1}(t), \\ e^{\lambda_j t}\mathbb{P}_{ij}(t) - \mathbb{P}_{ij}(0) &= \lambda_{j-1}\int_0^t e^{\lambda_j s}\mathbb{P}_{i,j-1}(s)\,\mathrm{d}s. \end{aligned}$$
Since $j > i$, we have $\mathbb{P}_{ij}(0) = 0$. Therefore
$$\mathbb{P}_{ij}(t) = \lambda_{j-1} e^{-\lambda_j t}\int_0^t e^{\lambda_j s}\mathbb{P}_{i,j-1}(s)\,\mathrm{d}s,$$
which is the recursive formula for $\mathbb{P}_{ij}(t)$ when $i < j$. Now consider the special case
$$\lambda_i = \lambda, \quad i = 0, 1, 2, \ldots \qquad \text{and} \qquad \mu_i = 0, \quad i = 1, 2, \ldots.$$
This is the Poisson process case from Case (3) above. We show that
$$\mathbb{P}_{i,i+k}(t) = e^{-\lambda t}\frac{(\lambda t)^k}{k!}, \quad k = 0, 1, 2, \ldots$$

by induction. For $k = 0$, we have $\mathbb{P}_{ii}(t) = e^{-\lambda t}$. Assume that for some $k \geq 0$,

$$\mathbb{P}_{i,i+k}(t) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}.$$

Using the recursive formula,

$$\begin{aligned}
\mathbb{P}_{i,i+k+1}(t) &= \lambda \int_0^t e^{-\lambda(t-s)} \mathbb{P}_{i,i+k}(s)\,\mathrm{d}s \\
&= \lambda e^{-\lambda t} \int_0^t e^{\lambda s} \frac{e^{-\lambda s}(\lambda s)^k}{k!}\,\mathrm{d}s \\
&= \frac{e^{-\lambda t}\lambda^{k+1}}{k!} \int_0^t s^k\,\mathrm{d}s \\
&= \frac{e^{-\lambda t}\lambda^{k+1}}{k!} \cdot \frac{t^{k+1}}{k+1} \\
&= e^{-\lambda t} \frac{(\lambda t)^{k+1}}{(k+1)!}.
\end{aligned}$$

Continuous-Time Recurrence and Stationarity

Recall that if the continuous-time chain $\{X(t)\}_{t\geq 0}$ is observed only when the state changes, the resulting process is a discrete-time Markov chain $\{J_n\}_{n\geq 0}$, called the discrete skeleton or embedded discrete-time Markov chain. For a nonabsorbing state i, its transition probabilities are

$$\mathbb{P}_{ij} = \mathbb{P}(J_{n+1} = j \mid J_n = i) = \frac{q_{ij}}{\nu_i}, \qquad j \neq i.$$

We now ask how the discrete-time notions translate into continuous time.

Lecture 18 — Monday, November 10, 2014

The special cases of birth and death processes include pure death processes, pure birth processes, Poisson processes, Yule processes, linear growth models, and $M/M/s$ queues. For a pure birth process, the forward equation gives the recursive formula

$$\mathbb{P}_{ij}(t) = \lambda_{j-1} e^{-\lambda_j t} \int_0^t e^{\lambda_j s} \mathbb{P}_{i,j-1}(s)\,\mathrm{d}s.$$

For special cases such as $\lambda_i = \lambda$ and $\mu_i = 0$, a closed-form solution can be found.

Once we have $\mathbf{P}(t) = \{\mathbb{P}_{ij}(t)\}$, the distribution of $X(t)$ can be expressed as in the discrete-time case:

$$\mathbb{P}(X(t) = j) = \sum_{i \in S} \mathbb{P}(X(t) = j \mid X(0) = i)\mathbb{P}(X(0) = i) = \sum_{i \in S} \alpha_i \mathbb{P}_{ij}(t) = (\alpha \mathbf{P}(t))_j.$$

That is, $\boldsymbol{\alpha}_t = \mathbf{P}(t)^\top \boldsymbol{\alpha}_0$.

The transition probability matrix is determined by $\mathbf{R}$. If $f : S \to \mathbb{R}$ is a function on the state space, then

$$\mathbb{E}\left[f(X(t))\right] = \sum_{j \in S} f(j)\mathbb{P}(X(t) = j) = \boldsymbol{\alpha}_t^\top \boldsymbol{f} = \boldsymbol{\alpha}_0^\top \mathbf{P}(t)\boldsymbol{f},$$

where $\boldsymbol{f} = (f(0), f(1), \ldots)^\top$ is viewed as a column vector. The matrix $\mathbf{P}(t)$ is called the **transition kernel**.

Definition 18.1 State j is **accessible** from state i, written $i \to j$, if

$$\mathbb{P}_{ij}(t) = \mathbb{P}(X(t) = j \mid X(0) = i) > 0$$

for some $t \geq 0$. States i and j **communicate**, written $i \leftrightarrow j$, if each is accessible from the other. A continuous-time Markov chain is **irreducible** if all states communicate with one another.

Assume $\nu_i > 0$ for every $i \in S$, so that there are no absorbing states created by zero holding rates. Then i and j communicate in continuous time if and only if they communicate in the discrete skeleton. Hence, under this assumption, the continuous-time Markov chain is irreducible if and only if its discrete skeleton is irreducible.

Definition 18.2 Given $X(0) = i$, let R_i be the elapsed time from the initial entrance into i until the next entrance into i after at least one jump, and set $R_i = \infty$ if that return never occurs. This exclusion of the initial sojourn is essential: otherwise right-continuity would make the first positive time at which $X(t) = i$ equal to 0. State i is *recurrent* if $\mathbb{P}(R_i < \infty) = 1$, and **transient** if $\mathbb{P}(R_i = \infty) > 0$. A recurrent state is **positive recurrent** if $\mathbb{E}[R_i] < \infty$ and **null recurrent** if $\mathbb{E}[R_i] = \infty$.

The continuous-time Markov chain revisits state i if and only if its discrete skeleton revisits state i. Therefore recurrence and transience agree for a continuous-time Markov chain and its discrete skeleton. There is no analogous notion of periodicity in continuous time, since exponential sojourn times can be arbitrarily small or arbitrarily large.

Example 18.3 Consider the discrete-time Markov chain with transition matrix

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 & \dots & \\ q & 0 & p & 0 & \dots \\ q & 0 & 0 & p & \dots \\ \vdots & \vdots & \vdots & \ddots & \end{pmatrix}.$$

Here $p, q > 0$ and $p + q = 1$. For this chain,

$$f_{00}^{(1)} = 0 \qquad \text{and} \qquad f_{00}^{(n)} = p^{n-2}q \quad \text{for } n \geq 2.$$

Hence

$$f_{00} = \sum_{n=2}^{\infty} f_{00}^{(n)} = q \sum_{n=2}^{\infty} p^{n-2} = \frac{q}{1-p} = 1.$$

Moreover, the expected number of transitions required to return to 0 is

$$\sum_{n=2}^{\infty} n f_{00}^{(n)} = \sum_{n=2}^{\infty} n p^{n-2} q = 2 \sum_{n=2}^{\infty} p^{n-2} q + \sum_{n=2}^{\infty} (n-2) p^{n-2} q < \infty.$$

Thus the embedded discrete-time chain is positive recurrent.

Now construct a continuous-time Markov chain whose discrete skeleton is this chain. Let W be the number of transitions required to return to 0. If $T_i \sim \text{Exp}(\nu_i)$ is the sojourn time in state i, then the return time to 0 is

$$R_0 = T_0 + T_1 + \cdots + T_{W-1}.$$

Therefore

$$\begin{aligned} \mathbb{E}[R_0] &= \mathbb{E}\left[\mathbb{E}\left[\sum_{i=0}^{W-1} T_i \mid W\right]\right] \\ &= \sum_{n=2}^{\infty} \mathbb{P}(W = n) \cdot \mathbb{E}\left[\sum_{i=0}^{n-1} T_i \mid W = n\right] \\ &= \sum_{n=2}^{\infty} \mathbb{P}(W = n) \cdot \sum_{i=0}^{n-1} \mathbb{E}[T_i] \end{aligned}$$

$$= \sum_{i=0}^{\infty} \mathbb{E}[T_i] \cdot \sum_{n=i+1}^{\infty} \mathbb{P}(W = n)$$
$$= \sum_{i=0}^{\infty} \mathbb{E}[T_i] \cdot \mathbb{P}(W \geq i+1)$$
$$= \frac{1}{\nu_0} + \frac{1}{\nu_1} + \sum_{i=2}^{\infty} \frac{p^{i-1}}{\nu_i}.$$

By choosing the holding rates $\{\nu_i\}_{i\geq 0}$ so that the sojourn times increase sufficiently quickly far from 0, this expectation can be made infinite; for example, taking $\nu_i = p^i$ for large i makes the final series diverge. Thus a continuous-time Markov chain and its discrete skeleton need not have the same positive recurrence or null recurrence behaviour.

Lecture 19 — Wednesday, November 12, 2014

As in discrete time, recurrence classification is a class property. In an irreducible continuous-time Markov chain, all states are positive recurrent, all states are null recurrent, or all states are transient.

Definition 19.1 Let $\{X(t)\}_{t\geq 0}$ be a continuous-time Markov chain. A column vector $\boldsymbol{\pi} = (\pi_0, \pi_1, \pi_2, \dots)^\top$ with $\pi_i \geq 0$ for all $i \in S$ is a **stationary distribution** if

1. $\mathbf{1}_{|S|}^\top \boldsymbol{\pi} = 1$;
2. $\boldsymbol{\pi} = \mathbf{P}(t)^\top \boldsymbol{\pi}$ for all $t \geq 0$.

Remark 19.2 This distribution is stationary for the same reason as in discrete time. If the initial distribution is $\boldsymbol{\alpha}_0 = \boldsymbol{\pi}$, then

$$\boldsymbol{\alpha}_t = \mathbf{P}(t)^\top \boldsymbol{\alpha}_0 = \mathbf{P}(t)^\top \boldsymbol{\pi} = \boldsymbol{\pi}.$$

Thus the distribution of $X(t)$ is $\boldsymbol{\pi}$ for every $t \geq 0$.

The stationary condition can be written in terms of the generator. The equivalence below is automatic in finite state spaces; in countably infinite spaces we use it under the standing regularity and summability conditions that justify differentiating and summing componentwise. If $\boldsymbol{\pi} =$

$\mathbf{P}(t)^\top \boldsymbol{\pi}$ for all $t \geq 0$, then

$$(\mathbf{P}(t)^\top - \mathbf{I})\boldsymbol{\pi} = \mathbf{0}_{|S|}.$$

Dividing by t and taking $t \to 0$ gives

$$\mathbf{R}^\top \boldsymbol{\pi} = \mathbf{0}_{|S|}.$$

Conversely, if $\mathbf{R}^\top \boldsymbol{\pi} = \mathbf{0}_{|S|}$ and $\boldsymbol{\alpha}_0 = \boldsymbol{\pi}$, then, using the backward equation,

$$(\boldsymbol{\alpha}_t)' = ((\mathbf{P}(t)^\top \boldsymbol{\alpha}_0))' = (\mathbf{P}'(t))^\top \boldsymbol{\alpha}_0 = \mathbf{P}(t)^\top \mathbf{R}^\top \boldsymbol{\pi} = \mathbf{0}_{|S|}.$$

Hence $\boldsymbol{\alpha}_t$ is constant in t, and $\boldsymbol{\pi}$ is stationary. Thus $\boldsymbol{\pi}$ is a stationary distribution if and only if

$$\mathbf{1}_{|S|}^\top \boldsymbol{\pi} = 1 \qquad \text{and} \qquad \mathbf{R}^\top \boldsymbol{\pi} = \mathbf{0}_{|S|}.$$

Assume $0 < \nu_i < \infty$ for every state. Comparing $\mathbf{R}^\top \boldsymbol{\pi} = \mathbf{0}_{|S|}$ with the invariant equations for the discrete skeleton gives a useful correspondence. The jth component of $\mathbf{R}^\top \boldsymbol{\pi} = \mathbf{0}_{|S|}$ is

$$\pi_j \nu_j = \sum_{i \neq j} \pi_i q_{ij}.$$

Since $q_{ij} = \nu_i \mathbb{P}_{ij}$, this becomes

$$\pi_j \nu_j = \sum_{i \neq j} \pi_i \nu_i \mathbb{P}_{ij}.$$

For the discrete skeleton, an invariant measure $\boldsymbol{\psi}$ satisfies $\psi_j = \sum_{i \in S} \psi_i \mathbb{P}_{ij}$, and since $\mathbb{P}_{ii} = 0$, we can write

$$\psi_j = \sum_{i \neq j} \psi_i \mathbb{P}_{ij}.$$

Thus $\boldsymbol{\pi} \odot \boldsymbol{\nu} = (\pi_0 \nu_0, \pi_1 \nu_1, \ldots)$ is an invariant measure for the skeleton whenever $\boldsymbol{\pi}$ is stationary for the continuous-time chain. Conversely, if $\boldsymbol{\psi}$ is stationary for the skeleton, then $(\psi_0/\nu_0, \psi_1/\nu_1, \ldots)$ satisfies the continuous-time invariant equations. These measures become probability distributions precisely when the relevant normalizing constants are finite:

$$\psi_j = \frac{\pi_j \nu_j}{\sum_{k \in S} \pi_k \nu_k} \qquad \text{and} \qquad \pi_j = \frac{\psi_j / \nu_j}{\sum_{k \in S} \psi_k / \nu_k}$$

and in this case, we have a one-to-one correspondence between the two stationary distributions.

The distributions $\boldsymbol{\psi}$ and $\boldsymbol{\pi}$ are not generally the same. The skeleton distribution ψ_j records the long-run proportion of visits to state j, while the continuous-time Markov chain distribution π_j records the long-run proportion of time spent in state j. Since a visit to j has mean holding time $1/\nu_j$, the continuous-time Markov chain stationary distribution weights skeleton visits by these mean holding times.

Remark 19.3

1. If the continuous-time Markov chain is irreducible and both $\boldsymbol{\pi}$ and $\boldsymbol{\psi}$ exist, then $\boldsymbol{\psi}$ is unique and hence $\boldsymbol{\pi}$ is unique.
2. It is possible for $\boldsymbol{\psi}$ to exist while $\boldsymbol{\pi}$ does not, because the denominator in
$$\pi_j = \frac{\psi_j/\nu_j}{\sum_{k\in S}\psi_k/\nu_k}$$
may diverge. The ν_i can be chosen arbitrarily; choosing them small enough makes normalization impossible. For example, if S is infinite, then taking $\nu_j = \psi_j$ gives $\psi_j/\nu_j = 1$ for every j, so the normalizing sum diverges.

From $\mathbf{R}^\top \boldsymbol{\pi} = \mathbf{0}_{|S|}$, we obtained
$$\pi_j\nu_j = \sum_{i\neq j}\pi_i q_{ij}, \quad j\in S. \tag{19.1}$$
Here ν_j can be thought of as the rate of leaving state j, given that the chain is in state j. On the other side, $\pi_i q_{ij}$ is the rate of transitioning from state i to state j. Thus, under stationarity, $\pi_j\nu_j$ is the "probability flow" leaving state j, while $\sum_{i\neq j}\pi_i q_{ij}$ is the total "probability flow" entering state j.

Intuitively, the "probability flow" leaving state j should equal the "probability flow" entering state j for every $j\in S$, as illustrated in Figure 10. This is the reason why the Markov chain is stationary. Because of this interpretation, (19.1) is sometimes called the **balance equation**.

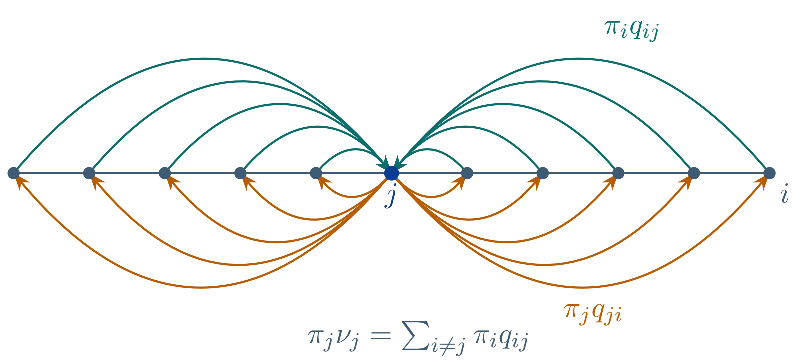


Figure 10

Lecture 20 — Monday, November 17, 2014

The relationship between the stationary distribution $\boldsymbol{\psi}$ of the discrete skeleton and the stationary distribution $\boldsymbol{\pi}$ of the continuous-time Markov chain is

$$\pi_j = \frac{\psi_j/\nu_j}{\sum_{k\in S}\psi_k/\nu_k} \qquad \text{and} \qquad \psi_j = \frac{\pi_j\nu_j}{\sum_{k\in S}\pi_k\nu_k},$$

provided the denominators are finite. This gives two ways to find the stationary distribution.

For an irreducible continuous-time Markov chain, π_j is also the long-run proportion of time the process spends in state j, and

$$\pi_j\nu_j = \sum_{i\neq j}\pi_i q_{ij}$$

are the balance equations. Here $\pi_j\nu_j$ is the "probability flow" leaving state j, and

$$\sum_{i\neq j}\pi_i q_{ij}$$

is the "probability flow" entering state j.

Theorem 20.1 (Basic limit theorem for continuous-time Markov chains) Let $\{X(t)\}_{t\geq 0}$ be an irreducible, nonexplosive continuous-time Markov chain with $0 < \nu_j < \infty$ for every state j. If the chain is recurrent, then

$$\lim_{t\to\infty}\mathbb{P}_{ij}(t) = \frac{\mathbb{E}[T_j]}{\mathbb{E}[R_j]} = \frac{1/\nu_j}{\mathbb{E}[R_j]},$$

where R_j is the return-cycle duration defined above. If the chain is positive recurrent, then the limiting vector is the unique stationary distribution $\boldsymbol{\pi}$, so

$$\lim_{t\to\infty}\mathbb{P}_{ij}(t) = \pi_j.$$

Remark 20.2 If the continuous-time Markov chain is positive recurrent, then $\mathbb{E}[R_j] < \infty$ and $\pi_j > 0$ for all $j \in S$. If the continuous-time Markov chain is null recurrent, then $\mathbb{E}[R_j] = \infty$ and no stationary distribution exists.

Partial proof. Any limiting distribution must be stationary. For any $t, s \geq 0$, the Chapman–Kolmogorov equations give

$$\mathbb{P}_{ij}(t+s) = \sum_{k \in S} \mathbb{P}_{ik}(t)\mathbb{P}_{kj}(s).$$

Suppose that these limits exist and that the limiting vector $\boldsymbol{\ell}$ has total mass 1. Pointwise convergence of the probability vectors to a probability vector then implies convergence in ℓ^1, so we may take $t \to \infty$ through the sum to obtain

$$\lim_{t\to\infty} \mathbb{P}_{ij}(t+s) = \sum_{k \in S} \left(\lim_{t\to\infty} \mathbb{P}_{ik}(t) \right) \mathbb{P}_{kj}(s).$$

Thus $\boldsymbol{\ell} = (\lim_{t\to\infty} \mathbb{P}_{i0}(t), \lim_{t\to\infty} \mathbb{P}_{i1}(t), \dots)^\top$ satisfies $\boldsymbol{\ell} = \mathbf{P}(s)^\top \boldsymbol{\ell}$ for every $s \geq 0$, and hence $\boldsymbol{\ell}$ is stationary. $\square$

Example 20.3 Let

$$\mathbf{R} = \begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix}, \quad \alpha, \beta > 0.$$

The chain is irreducible and positive recurrent. Solving $\mathbf{R}^\top \boldsymbol{\pi} = \mathbf{0}_2$ and $\mathbf{1}_2^\top \boldsymbol{\pi} = 1$ gives

$$\pi_0 = \frac{\beta}{\alpha+\beta} \qquad \text{and} \qquad \pi_1 = \frac{\alpha}{\alpha+\beta}.$$

Thus

$$\boldsymbol{\pi} = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta} \right)$$

is the unique stationary distribution.

The transition matrix is

$$\mathbf{P}(t) = \begin{pmatrix} \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta} e^{-(\alpha+\beta)t} & \frac{\alpha}{\alpha+\beta} - \frac{\alpha}{\alpha+\beta} e^{-(\alpha+\beta)t} \\ \frac{\beta}{\alpha+\beta} - \frac{\beta}{\alpha+\beta} e^{-(\alpha+\beta)t} & \frac{\alpha}{\alpha+\beta} + \frac{\beta}{\alpha+\beta} e^{-(\alpha+\beta)t} \end{pmatrix}.$$

Therefore

$$\lim_{t\to\infty} \mathbf{P}(t) = \begin{pmatrix} \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \\ \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \end{pmatrix}.$$

The limiting probability does not depend on the initial state, and each row is the unique stationary distribution.

Also, $T_0 \sim \text{Exp}(\alpha)$ and $T_1 \sim \text{Exp}(\beta)$, so

$$\mathbb{E}[R_0] = \mathbb{E}[T_0] + \mathbb{E}[T_1] = \frac{1}{\alpha} + \frac{1}{\beta} = \frac{\alpha+\beta}{\alpha\beta}.$$

Hence

$$\frac{1/\nu_0}{\mathbb{E}[R_0]} = \frac{1/\alpha}{(\alpha+\beta)/(\alpha\beta)} = \frac{\beta}{\alpha+\beta} = \pi_0,$$

as predicted by the basic limit theorem.

Remark 20.4 For an irreducible continuous-time Markov chain, a stationary distribution exists if and only if the chain is positive recurrent.

For a birth and death process, recall that

$$\mathbf{R} = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & \cdots \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & \cdots \\ 0 & \mu_2 & -(\lambda_2+\mu_2) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

The balance equations $\mathbf{R}^\top \boldsymbol{\pi} = \mathbf{0}_{|S|}$ equate the rate out of each state with the rate into that state:

$$\pi_0\lambda_0 = \pi_1\mu_1 \qquad \text{and} \qquad \pi_1(\mu_1+\lambda_1) = \pi_0\lambda_0 + \pi_2\mu_2.$$

Using the first equation in the second gives

$$\pi_1\lambda_1 = \pi_2\mu_2.$$

Continuing inductively,

$$\pi_n\lambda_n = \pi_{n+1}\mu_{n+1} \qquad \text{and hence} \qquad \pi_n = \pi_0 \prod_{i=0}^{n-1} \frac{\lambda_i}{\mu_{i+1}}.$$

Normalizing gives

$$\pi_0 = \left(1 + \sum_{r=1}^{\infty} \prod_{i=0}^{r-1} \frac{\lambda_i}{\mu_{i+1}}\right)^{-1},$$

provided the sum is finite. Hence

$$\pi_n = \left(1 + \sum_{r=1}^{\infty} \prod_{i=0}^{r-1} \frac{\lambda_i}{\mu_{i+1}}\right)^{-1} \prod_{i=0}^{n-1} \frac{\lambda_i}{\mu_{i+1}}.$$

Thus a stationary distribution exists if and only if

$$\sum_{n=0}^{\infty} \prod_{i=0}^{n-1} \frac{\lambda_i}{\mu_{i+1}} < \infty,$$

in which case the stationary distribution and the limiting distribution coincide. If the normalizing constant is infinite, no probability distribution can be obtained from these invariant weights, so the irreducible chain is either null recurrent or transient. We conclude the following:

Theorem 20.5 An irreducible, nonexplosive birth and death process is positive recurrent if and only if

$$\sum_{n=0}^{\infty} \prod_{i=0}^{n-1} \frac{\lambda_i}{\mu_{i+1}} < \infty.$$

Lecture 21 — Wednesday, November 19, 2014

The stationary distribution $\boldsymbol{\pi}$ exists if and only if the Markov chain has at least one positive recurrent class, and it is unique if and only if the Markov chain has *exactly* one positive recurrent class.

Question 21.1 If the condition

$$\sum_{n=0}^{\infty} \prod_{i=0}^{n-1} \frac{\lambda_i}{\mu_{i+1}} < \infty$$

is not satisfied, how can we determine whether the birth and death process is null recurrent or transient?

Consider an irreducible birth and death process, so that $\lambda_j, \mu_j > 0$ for the relevant states. Its discrete skeleton has transition matrix

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 & \cdots \\ \dfrac{\mu_1}{\lambda_1+\mu_1} & 0 & \dfrac{\lambda_1}{\lambda_1+\mu_1} & \cdots \\ 0 & \dfrac{\mu_2}{\lambda_2+\mu_2} & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Recall that

$$f_{n0} = \mathbb{P}(\text{the chain ever visits state } 0 \mid X_0 = n).$$

Then $f_{00} = 1$ if and only if the irreducible chain is recurrent, while $f_{00} < 1$ if and only if it is transient.

For a first-step analysis, we have

$$\begin{aligned} f_{00} &= f_{10}, \\ f_{10} &= \frac{\mu_1}{\lambda_1+\mu_1} + \frac{\lambda_1}{\lambda_1+\mu_1} f_{20}. \end{aligned}$$

Hence

$$(\lambda_1+\mu_1) f_{10} = \mu_1 + \lambda_1 f_{20},$$

so

$$\lambda_1(f_{20} - f_{10}) = \mu_1(f_{10} - 1) \qquad \text{and} \qquad f_{20} - f_{10} = \frac{\mu_1}{\lambda_1}(f_{10} - 1).$$

In general, for $n \geq 1$,

$$f_{n0} = \frac{\mu_n}{\lambda_n+\mu_n} f_{n-1,0} + \frac{\lambda_n}{\lambda_n+\mu_n} f_{n+1,0},$$

which is equivalent to

$$\lambda_n(f_{n+1,0} - f_{n0}) = \mu_n(f_{n0} - f_{n-1,0}).$$

Define

$$\Delta_n := f_{n+1,0} - f_{n0}.$$

Then

$$\Delta_n = \frac{\mu_n}{\lambda_n}\Delta_{n-1} \qquad \text{and} \qquad \Delta_0 = f_{10} - 1.$$

By iteration,

$$\Delta_n = \left(\prod_{i=1}^{n} \frac{\mu_i}{\lambda_i}\right)(f_{10} - 1),$$

that is,

$$f_{n0} - f_{n-1,0} = \left(\prod_{i=1}^{n} \frac{\mu_i}{\lambda_i}\right)(f_{10} - 1), \quad n \geq 1.$$

Now telescope:

$$f_{n+1,0} = f_{10} + \sum_{k=1}^{n}(f_{k+1,0} - f_{k0}) = f_{10} + (f_{10} - 1)\sum_{k=1}^{n}\prod_{i=1}^{k} \frac{\mu_i}{\lambda_i}.$$

Since probabilities are nonnegative, $f_{n+1,0} \geq 0$ for every n. If the chain is transient, then $f_{00} < 1$, and because $f_{00} = f_{10}$ we have $f_{10} - 1 < 0$. Therefore

$$\sum_{k=1}^{n}\prod_{i=1}^{k} \frac{\mu_i}{\lambda_i} \leq \frac{f_{10}}{1 - f_{10}} = \frac{f_{00}}{1 - f_{00}} < \infty.$$

Taking $n \to \infty$ yields the necessary condition that the series

$$\sum_{k=1}^{\infty} \frac{\mu_k \cdots \mu_1}{\lambda_k \cdots \lambda_1} = \sum_{k=1}^{\infty}\prod_{i=1}^{k} \frac{\mu_i}{\lambda_i}.$$

converges. For the converse and the exact hitting probability, stop the skeleton when it reaches either 0 or a finite level $N > 1$. Define

$$a_0 = 1 \qquad \text{and} \qquad a_k = \prod_{i=1}^{k} \frac{\mu_i}{\lambda_i}, \qquad k \geq 1.$$

Solving the same difference recurrence with boundary values $h_0^{(N)} = 1$ and $h_N^{(N)} = 0$ gives

$$h_n^{(N)} = \mathbb{P}_n(T_0 < T_N) = 1 - \frac{\sum_{k=0}^{n-1} a_k}{\sum_{k=0}^{N-1} a_k}, \qquad 0 \leq n \leq N.$$

For $n = 1$, the events $\{T_0 < T_N\}$ increase to $\{T_0 < \infty\}$, so

$$f_{10} = 1 - \frac{1}{\sum_{k=0}^{\infty} a_k}.$$

Thus $f_{00} = f_{10} < 1$ exactly when $\sum_{k \geq 1} a_k < \infty$. Hence the chain is transient if and only if

$$\sum_{k=1}^{\infty}\prod_{i=1}^{k} \frac{\mu_i}{\lambda_i} < \infty.$$

If the positive recurrence condition fails and this transience condition also fails, then the chain is null recurrent.

Example 21.2 For the $M/M/1$ queue,

$$\lambda_n = \lambda, \quad n = 0, 1, 2, \ldots \qquad \text{and} \qquad \mu_n = \mu, \quad n = 1, 2, \ldots.$$

The transience criterion becomes

$$\sum_{k=1}^{\infty} \left(\frac{\mu}{\lambda}\right)^k < \infty,$$

which holds if and only if $\mu < \lambda$. Meanwhile, the positive recurrence criterion gives positive recurrence if $\lambda < \mu$. Therefore the $M/M/1$ queue is positive recurrent if $\lambda < \mu$, transient if $\lambda > \mu$, and null recurrent if $\lambda = \mu$.

Continuous Phase-Type Distributions

Let $\{X(t)\}_{t\geq 0}$ be a continuous-time Markov chain with transient states

$$E = \{1, 2, \ldots, m\}$$

and a single absorbing state 0. With the absorbing state listed first, the generator has block form

$$\mathbf{R} = \begin{pmatrix} 0 & \mathbf{0}_{1\times m} \\ \boldsymbol{t}^0 & \mathbf{T} \end{pmatrix},$$

where $\mathbf{T}$ is the $m \times m$ transient block and $\boldsymbol{t}^0$ is the column vector of rates from the transient states to the absorbing state. Since the rows of $\mathbf{R}$ sum to zero,

$$\boldsymbol{t}^0 + \mathbf{T}\mathbf{1}_m = 0 \implies \boldsymbol{t}^0 = -\mathbf{T}\mathbf{1}_m.$$

Let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)^\top$ be the initial probability column vector over the transient states. Then

$$\alpha_i = \mathbb{P}(X(0) = i) \qquad \text{and} \qquad \alpha_0 = \mathbb{P}(X(0) = 0) = 1 - \sum_{i=1}^{m} \alpha_i = 1 - \boldsymbol{\alpha}^\top \mathbf{1}_m.$$

Define the absorption time

$$Y = \min\{t \geq 0 : X(t) = 0\}.$$

The mechanism behind this definition is shown in Figure 11: the chain moves among the transient states according to $\mathbf{T}$ until an exit governed by $\boldsymbol{t}^0$ sends it to state 0.

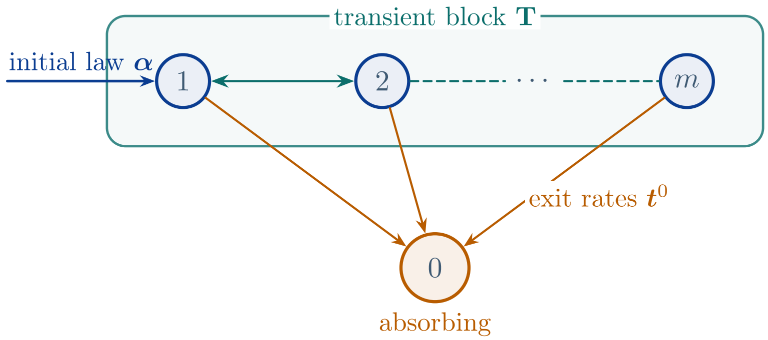


FIGURE 11

For $y \geq 0$,

$$\begin{aligned}
\mathbb{P}(Y > y) &= \mathbb{P}(X(y) \in E) \\
&= \sum_{i=1}^{m} \alpha_i \mathbb{P}(X(y) \in E \mid X(0) = i) \\
&= \sum_{i=1}^{m} \alpha_i \sum_{j=1}^{m} \mathbb{P}(X(y) = j \mid X(0) = i) \\
&= \sum_{i=1}^{m} \alpha_i \sum_{j=1}^{m} \mathbb{P}_{ij}(y).
\end{aligned}$$

The transition matrix at time y is

$$\mathbf{P}(y) = e^{\mathbf{R}y} = \mathbf{I} + \sum_{n=1}^{\infty} \frac{y^n}{n!} \mathbf{R}^n.$$

Using the block structure of $\mathbf{R}$,

$$\mathbf{R}^n = \begin{pmatrix} 0 & \mathbf{0}_{1\times m} \\ \mathbf{T}^{n-1}\boldsymbol{t}^0 & \mathbf{T}^n \end{pmatrix},$$

so

$$\mathbf{P}(y) = \mathbf{I} + \sum_{n=1}^{\infty} \frac{y^n}{n!} \begin{pmatrix} 0 & \mathbf{0}_{1\times m} \\ \mathbf{T}^{n-1}\boldsymbol{t}^0 & \mathbf{T}^n \end{pmatrix}$$

and the transient-to-transient block of $\mathbf{P}(y)$ is $e^{\mathbf{T}y}$. Because the rows sum to 1, the full matrix has the form

$$\mathbf{P}(y) = \begin{pmatrix} 1 & \mathbf{0}_{1\times m} \\ \mathbf{1}_m - e^{\mathbf{T}y}\mathbf{1}_m & e^{\mathbf{T}y} \end{pmatrix}.$$

Consequently,

$$\mathbb{P}(Y > y) = \sum_{i=1}^{m} \alpha_i \sum_{j=1}^{m} (e^{\mathbf{T}y})_{ij} = \boldsymbol{\alpha}^\top e^{\mathbf{T}y} \mathbf{1}_m$$

and so the cumulative distribution function of Y is

$$F_Y(y) = \mathbb{P}(Y \le y) = 1 - \boldsymbol{\alpha}^\top e^{\mathbf{T}y} \mathbf{1}_m.$$

In particular,

$$\mathbb{P}(Y = 0) = F_Y(0) = 1 - \boldsymbol{\alpha}^\top \mathbf{1}_m = \alpha_0.$$

For $y > 0$, the density is

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) = -\boldsymbol{\alpha}^\top \mathbf{T} e^{\mathbf{T}y} \mathbf{1}_m = \boldsymbol{\alpha}^\top e^{\mathbf{T}y} \boldsymbol{t}^0.$$

Thus Y has a point mass α_0 at 0 and density $\boldsymbol{\alpha}^\top e^{\mathbf{T}y} \boldsymbol{t}^0$ for $y > 0$.

Lecture 22 — Monday, November 24, 2014

Definition 22.1 A random variable Y constructed as the absorption time above is said to have a **continuous phase-type distribution** with representation $(\boldsymbol{\alpha}, \mathbf{T})$. We write

$$Y \sim \mathrm{CPH}_m(\boldsymbol{\alpha}, \mathbf{T}).$$

Here $\boldsymbol{\alpha}$ is the initial subdistribution over the m transient phases, with $\boldsymbol{\alpha} \ge \mathbf{0}$ and $\boldsymbol{\alpha}^\top \mathbf{1}_m \le 1$, and $\mathbf{T}$ is the transient part of the generator matrix. The missing mass is $\mathbb{P}(Y = 0)$.

Example 22.2 The exponential distribution is a continuous phase-type distribution. If $Y \sim \mathrm{Exp}(\lambda)$, take one transient state and one absorbing state, with

$$\mathbf{R} = \begin{pmatrix} 0 & 0 \\ \lambda & -\lambda \end{pmatrix}, \qquad \boldsymbol{\alpha} = (1), \qquad \mathbf{T} = (-\lambda), \qquad \text{and} \qquad \boldsymbol{t}^0 = (\lambda).$$

Then $Y \sim \mathrm{CPH}_1(1, -\lambda)$.

We list out some properties of continuous phase-type distributions.

1. *Independent sums.* If $X \sim \mathrm{CPH}_m(\boldsymbol{\alpha}, \mathbf{T})$ and $Y \sim \mathrm{CPH}_n(\boldsymbol{\beta}, \mathbf{S})$ are independent, then

$$X + Y \sim \mathrm{CPH}_{m+n}(\boldsymbol{\delta}, \mathbf{D}),$$

where $\alpha_0 = 1 - \boldsymbol{\alpha}^\top \mathbf{1}_m = \mathbb{P}(X = 0)$,

$$\boldsymbol{\delta} = \begin{pmatrix} \boldsymbol{\alpha} \\ \alpha_0 \boldsymbol{\beta} \end{pmatrix} \qquad \text{and} \qquad \mathbf{D} = \begin{pmatrix} \mathbf{T} & \boldsymbol{t}^0 \boldsymbol{\beta}^\top \\ \mathbf{0}_{n\times m} & \mathbf{S} \end{pmatrix}.$$

Thus continuous phase-type distributions are closed under independent sums.

As a special case, the **Erlang distribution** with density

$$f(x) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!}, \quad x > 0,$$

is continuous phase-type, since an Erlang(n, λ) random variable is the sum of n independent Exp(λ) random variables. One representation uses

$$\boldsymbol{\alpha} = (1, 0, \dots, 0) \qquad \text{and} \qquad \mathbf{T} = \begin{pmatrix} -\lambda & \lambda & 0 & \cdots \\ 0 & -\lambda & \lambda & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

The same construction works for independent sums of exponential random variables with different parameters.

2. *Finite mixtures.* If $X \sim \mathrm{CPH}_m(\boldsymbol{\alpha}, \mathbf{T})$ and $Y \sim \mathrm{CPH}_n(\boldsymbol{\beta}, \mathbf{S})$, and

$$Z = \begin{cases} X, & \text{with probability } p, \\ Y, & \text{with probability } 1-p, \end{cases}$$

where the choice is independent of X and Y. Then $F_Z = pF_X + (1-p)F_Y$ and $Z \sim \mathrm{CPH}_{m+n}(\boldsymbol{\gamma}, \mathbf{G})$, where

$$\boldsymbol{\gamma} = \begin{pmatrix} p\boldsymbol{\alpha} \\ (1-p)\boldsymbol{\beta} \end{pmatrix} \qquad \text{and} \qquad \mathbf{G} = \begin{pmatrix} \mathbf{T} & \mathbf{0}_{m\times n} \\ \mathbf{0}_{n\times m} & \mathbf{S} \end{pmatrix}.$$

Thus mixtures of exponential or Erlang distributions are again continuous phase-type. For example,

$$f(x) = p\frac{\lambda^2 x e^{-\lambda x}}{1!} + (1-p)\frac{\lambda^7 x^6 e^{-\lambda x}}{6!}, \quad x > 0,$$

is a mixture of Erlang$(2, \lambda)$ and Erlang$(7, \lambda)$ distributions, hence is continuous phase-type.

A direct block construction gives a nine-phase representation:

$$Z \sim \mathrm{CPH}_9(\widetilde{\boldsymbol{\alpha}}, \widetilde{\mathbf{T}}) \qquad \text{and} \qquad \widetilde{\boldsymbol{\alpha}} = \begin{pmatrix} p \\ 0 \\ 1-p \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

with

$$\widetilde{\mathbf{T}} = \begin{pmatrix} \mathbf{T}_2 & \mathbf{0}_{2\times 7} \\ \mathbf{0}_{7\times 2} & \mathbf{T}_7 \end{pmatrix},$$

where $\mathbf{T}_r$ is the $r \times r$ Erlang transient block with $-\lambda$ on the diagonal and λ on the super-diagonal.

We can also obtain a smaller representation. Let

$$\mathbf{T}_7 = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & 0 & 0 & 0 \\ 0 & -\lambda & \lambda & 0 & 0 & 0 & 0 \\ 0 & 0 & -\lambda & \lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & -\lambda & \lambda & 0 & 0 \\ 0 & 0 & 0 & 0 & -\lambda & \lambda & 0 \\ 0 & 0 & 0 & 0 & 0 & -\lambda & \lambda \\ 0 & 0 & 0 & 0 & 0 & 0 & -\lambda \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{\beta} = (1-p, 0, 0, 0, 0, p, 0).$$

Starting in phase 1 gives Erlang$(7, \lambda)$, while starting in phase 6 gives Erlang$(2, \lambda)$. Therefore

$$Z \sim \mathrm{CPH}_7(\boldsymbol{\beta}, \mathbf{T}_7).$$

This shows explicitly that continuous phase-type representations are not unique.

3. *Finite moments.* If $X \sim \mathrm{CPH}_m(\boldsymbol{\alpha}, \mathbf{T})$, then X has all finite moments. Let

$$\mathcal{L}(s) = \mathbb{E}[e^{-sX}]$$

be its Laplace transform. Since X has point mass α_0 at 0 and density $\boldsymbol{\alpha}^\top e^{\mathbf{T}x}\boldsymbol{t}^0$ for $x > 0$,

$$\mathcal{L}(s) = \alpha_0 + \int_0^\infty e^{-sx}\boldsymbol{\alpha}^\top e^{\mathbf{T}x}\boldsymbol{t}^0 \,\mathrm{d}x$$

$$
\begin{aligned}
&= \alpha_0 + \boldsymbol{\alpha}^\top \int_0^\infty e^{(\mathbf{T}-s\mathbf{I})x}\,\mathrm{d}x\,\boldsymbol{t}^0 \\
&= \alpha_0 + \boldsymbol{\alpha}^\top \left[e^{(\mathbf{T}-s\mathbf{I})x}(\mathbf{T}-s\mathbf{I})^{-1} \right]_{x=0}^{\infty} \boldsymbol{t}^0 \\
&= \alpha_0 + \boldsymbol{\alpha}^\top (s\mathbf{I}-\mathbf{T})^{-1}\boldsymbol{t}^0
\end{aligned}
$$

for $s \geq 0$. The boundary term at ∞ is 0 because all eigenvalues of $\mathbf{T} - s\mathbf{I}$ have strictly negative real part for $s \geq 0$.

The eigenvalues of the finite transient generator $\mathbf{T}$ have strictly negative real parts, so $x^n e^{\mathbf{T}x} \to \mathbf{0}$ as $x \to \infty$ for every n. We can therefore calculate the moments directly from the density. Put

$$\mathbf{I}_n = \int_0^\infty x^n e^{\mathbf{T}x}\,\mathrm{d}x.$$

Since $\frac{\mathrm{d}}{\mathrm{d}x} e^{\mathbf{T}x} = \mathbf{T}e^{\mathbf{T}x}$, integration by parts gives

$$\mathbf{I}_0 = (-\mathbf{T})^{-1} \qquad \text{and} \qquad \mathbf{I}_n = n(-\mathbf{T})^{-1}\mathbf{I}_{n-1} = n!(-\mathbf{T})^{-n-1}.$$

Using $\boldsymbol{t}^0 = -\mathbf{T}\mathbf{1}_m$, we obtain

$$
\begin{aligned}
\mathbb{E}[X^n] &= \boldsymbol{\alpha}^\top \mathbf{I}_n \boldsymbol{t}^0 \\
&= n!\,\boldsymbol{\alpha}^\top (-\mathbf{T})^{-n-1}(-\mathbf{T})\mathbf{1}_m \\
&= n!\,\boldsymbol{\alpha}^\top (-\mathbf{T})^{-n}\mathbf{1}_m, \quad n = 1, 2, 3, \ldots.
\end{aligned}
$$

Hence every moment is finite. In particular,

$$\mathbb{E}[X] = \boldsymbol{\alpha}^\top (-\mathbf{T})^{-1}\mathbf{1}_m.$$

Remark 22.3 Continuous phase-type distributions are weakly dense in the set of probability distributions on $[0, \infty)$: every such distribution can be approximated in distribution by continuous phase-type laws.

Lecture 23 — Wednesday, November 26, 2014

4. Martingales

Definitions and Examples

Definition 23.1 A stochastic process $\{X_n\}_{n\geq 0}$ is a **martingale** if

1. $\mathbb{E}[|X_n|] < \infty$ for all n;
2. $\mathbb{E}[X_{n+1} \mid X_0, \ldots, X_n] = X_n$ for all n.

The guiding intuition is that a martingale models a fair game: conditional on the present and the past, the expected future value equals the current value.

Example 23.2 Let $S_0 = 0$ and let $X_1, X_2, \ldots$ be independent random variables with

$$\mathbb{P}(X_i = 1) = \frac{1}{2} = \mathbb{P}(X_i = -1).$$

The symmetric random walk $\{S_n\}_{n\geq 0}$ is defined by

$$S_n = \sum_{i=1}^{n} X_i.$$

Then $\{S_n\}_{n\geq 0}$ is a martingale. Indeed, $|S_n| \leq n$, so $\mathbb{E}[|S_n|] < \infty$, and

$$\mathbb{E}[S_{n+1} \mid S_0, \ldots, S_n] = \mathbb{E}[S_n + X_{n+1} \mid S_0, \ldots, S_n] = S_n + \mathbb{E}[X_{n+1}] = S_n.$$

Example 23.3 Let $X_1, X_2, \ldots$ be independent random variables with $\mathbb{E}[X_i] = 1$ and $\mathbb{E}[|X_i|] < \infty$. Define

$$Z_n = \prod_{i=1}^{n} X_i.$$

Set $Z_0 = 1$, the empty product. Independence gives $\mathbb{E}[|Z_n|] = \prod_{i=1}^{n} \mathbb{E}[|X_i|] < \infty$, and the process $\{Z_n\}_{n\geq 0}$ is a martingale, since

$$\mathbb{E}[Z_{n+1} \mid Z_0, \ldots, Z_n] = \mathbb{E}[Z_n X_{n+1} \mid Z_0, \ldots, Z_n] = Z_n \mathbb{E}[X_{n+1}] = Z_n.$$

Proposition 23.4 If $\{X_n\}_{n\geq 0}$ is a martingale, then $\mathbb{E}[X_n] = \mathbb{E}[X_0]$ for every n.

Proof. By iterating conditional expectation,

$$\mathbb{E}[X_n] = \mathbb{E}[\mathbb{E}[X_n \mid X_0, \ldots, X_{n-1}]] = \mathbb{E}[X_{n-1}] = \cdots = \mathbb{E}[X_0].$$

□

Proposition 23.5 If $\{X_n\}_{n\geq 0}$ is a martingale with finite variances, then

$$\mathrm{Var}(X_{n+1}) \geq \mathrm{Var}(X_n)$$

for every n.

Proof. Let $Y = X_{n+1} - X_n$. Since

$$\mathbb{E}[Y \mid X_0, \ldots, X_n] = 0,$$

we have

$$\mathrm{Cov}(Y, X_n) = \mathbb{E}[Y X_n] - \mathbb{E}[Y]\mathbb{E}[X_n] = \mathbb{E}[\mathbb{E}[Y X_n \mid X_0, \ldots, X_n]] = \mathbb{E}[X_n \mathbb{E}[Y \mid X_0, \ldots, X_n]] = 0.$$

Therefore

$$\mathrm{Var}(X_{n+1}) = \mathrm{Var}(X_n + Y) = \mathrm{Var}(X_n) + \mathrm{Var}(Y) \geq \mathrm{Var}(X_n).$$

□

Stopping Times and Optional Sampling

For deterministic times n, a martingale satisfies $\mathbb{E}[X_n] = \mathbb{E}[X_0]$. For random times T, this conclusion can fail.

Example 23.6 Let $\{S_n\}$ be the simple symmetric random walk and define

$$T = \min\left\{0 \le n \le N : S_n = \max_{0\le i\le N} S_i\right\}.$$

This is the first time the walk reaches its maximum over the full interval $\{0, \dots, N\}$. Then $S_T \ge 0$ and $\mathbb{P}(S_T > 0) > 0$, so

$$\mathbb{E}[S_T] > 0 = \mathbb{E}[S_0].$$

The issue is that T depends on future information.

Definition 23.7 A nonnegative integer-valued random variable T is a **stopping time** for the stochastic process $\{X_n\}_{n\ge 0}$ if the event $\{T = n\}$ is determined by $X_0, \dots, X_n$ for every n. Equivalently, for each n there is a function f_n such that

$$\mathbb{1}_{\{T=n\}} = f_n(X_0, \dots, X_n).$$

Intuitively, if T is a stopping time, then we know that the process has arrived when it arrives. That clears up everything.

Example 23.8 The first hitting time of a level h,

$$T_h = \min\{n : X_n = h\},$$

with the convention $T_h = \infty$ if the level is never reached, is a stopping time. By contrast, the last time the process hits a level h before time N,

$$T_h^e = \max\{n \le N : X_n = h\},$$

is generally not a stopping time, since deciding whether a time is the last visit requires future information.

The information distinction is visible in Figure 12. The first visit is recognizable when it occurs; a visit can be certified as the last one before N only after the remaining path has been observed.

Stopping times alone are not enough to say very much. For the simple symmetric random walk, the hitting time

$$T_1 = \min\{n : S_n = 1\}$$

is a stopping time, but

$$\mathbb{E}[S_{T_1}] = 1 \ne 0 = \mathbb{E}[S_0].$$

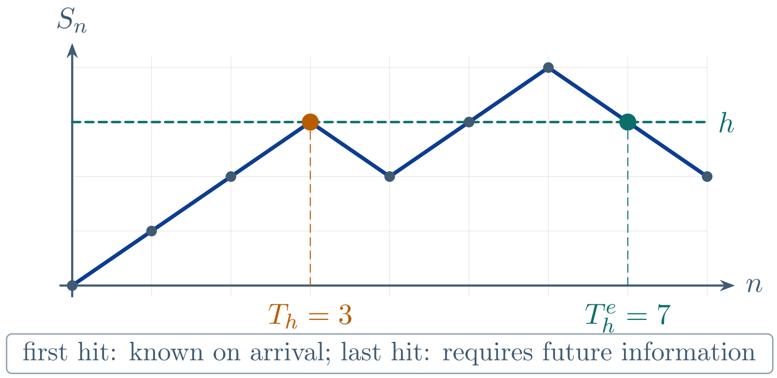


FIGURE 12

Additional hypotheses are needed.

Theorem 23.9 (Optional sampling) Let $\{X_n\}_{n\geq 0}$ be a martingale, and let T be a stopping time with $\mathbb{P}(T < \infty) = 1$. If

$$\mathbb{E}[|X_T|] < \infty$$

and

$$\lim_{n\to\infty} \mathbb{E}\big[|X_n|\mathbb{1}_{\{T>n\}}\big] = 0,$$

then

$$\mathbb{E}[X_T] = \mathbb{E}[X_0].$$

In particular, the conclusion holds when T is bounded. It also holds when the martingale $\{X_n\}$ is **uniformly integrable**:

$$\lim_{K\to\infty} \sup_n \mathbb{E}\big[|X_n|\mathbb{1}_{\{|X_n|>K\}}\big] = 0.$$

Example 23.10 (Gambler's ruin: symmetric case) Suppose $X_0 = m$, and from each interior state the gambler moves up or down by one with probabilities $p = q = 1/2$. The game ends when the wealth reaches 0 or N. Let

$$T = \min\{n : X_n = 0 \text{ or } X_n = N\}.$$

Then $\{X_n\}$ is a martingale, X_T is bounded, and the optional sampling theorem gives

$$\mathbb{E}[X_T] = \mathbb{E}[X_0] = m.$$

Since

$$\mathbb{E}[X_T] = 0 \cdot \mathbb{P}(X_T = 0) + N \cdot \mathbb{P}(X_T = N),$$

we obtain

$$\mathbb{P}(X_T = N) = \frac{m}{N} \qquad \text{and} \qquad \mathbb{P}(X_T = 0) = 1 - \frac{m}{N}.$$

Here $T < \infty$ almost surely because the finite absorbing chain reaches a boundary almost surely; moreover, on $\{T > n\}$ we have $0 < X_n < N$, so the tail condition in the theorem follows from $\mathbb{P}(T > n) \to 0$.

Lecture 24 — Monday, December 01, 2014

We continue our foray into martingales.

Example 24.1 (Gambler's ruin: asymmetric case) Now suppose $0 < p, q < 1$ and $p \neq q$, where $q = 1 - p$. Put

$$r = \frac{q}{p} \qquad \text{and} \qquad Y_n = r^{X_n}.$$

Then $\{Y_n\}$ is a martingale, since, conditional on $X_n = x$,

$$\mathbb{E}[Y_{n+1} \mid X_n = x] = pr^{x+1} + qr^{x-1} = r^x \left(pr + \frac{q}{r}\right) = r^x(q+p) = r^x.$$

Alternatively, write $Y_n = Y_{n-1}Z_n$, where

$$Z_n = \begin{cases} q/p & \text{if } X_n = X_{n-1} + 1 \\ p/q & \text{if } X_n = X_{n-1} - 1 \end{cases}$$

and the Z_n are all independent with $\mathbb{E}[Z_n] = (q/p)p + (p/q)q = 1$. By Example 23.3,

$$Y_n = Y_0 \prod_{i=1}^n Z_n = \left(\frac{q}{p}\right)^m \prod_{i=1}^n Z_n$$

is a martingale.

With the same stopping time

$$T = \min\{n : X_n = 0 \text{ or } X_n = N\},$$

both Y_T and every Y_n are bounded by $M = \max\{1, r^N\}$. Since $T < \infty$ almost surely,

$$\mathbb{E}\left[|Y_n| \mathbb{1}_{\{T>n\}}\right] \leq M\mathbb{P}(T > n) \longrightarrow 0.$$

The optional sampling theorem therefore gives

$$\mathbb{E}[Y_T] = \mathbb{E}[Y_0] = r^m.$$

If $a = \mathbb{P}(X_T = 0)$ is the probability of ruin, then

$$r^m = \mathbb{E}[r^{X_T}] = a + r^N(1-a).$$

Solving for a gives

$$\mathbb{P}(X_T = 0) = \frac{r^N - r^m}{r^N - 1} = \frac{(q/p)^N - (q/p)^m}{(q/p)^N - 1}.$$

This is substantially easier than the old-school way of deriving the ruin probability!

Martingale Convergence

Theorem 24.2 (Martingale convergence theorem) Let $\{X_n\}_{n\geq 0}$ be a martingale. If there exists $M < \infty$ such that

$$\sup_{n\geq 1} \mathbb{E}[|X_n|] \leq M$$

then $\lim_{n\to\infty} X_n$ exists and is finite with probability 1.

The condition $\sup_{n\geq 1} \mathbb{E}[|X_n|] \leq M < \infty$ means that $\{X_n\}$ is bounded in L^1.

We omit the proof because it requires measure theory beyond the scope of these notes.

Corollary 24.3 If $\{X_n\}_{n\geq 0}$ is a nonnegative martingale, then $\lim_{n\to\infty} X_n$ exists and is finite with probability 1.

Proof. Since $X_n \geq 0$,

$$\mathbb{E}[|X_n|] = \mathbb{E}[X_n] = \mathbb{E}[X_0]$$

for every n. Thus the martingale is uniformly bounded in L^1, and the martingale convergence theorem applies. $\square$

Example 24.4 Consider a discrete stock price model with $X_0 = a > 0$ and

$$X_n = a \prod_{i=1}^{n} Z_i \qquad \text{and} \qquad \mathbb{P}(Z_i = 1.2) = \mathbb{P}(Z_i = 0.8) = \frac{1}{2},$$

where $Z_1, Z_2, \ldots$ are independent. Each step is an upward or downward move of 20%.

First, $\{X_n\}$ is a positive martingale:

$$\mathbb{E}[X_{n+1} \mid X_0, \ldots, X_n] = \mathbb{E}[X_n Z_{n+1} \mid X_0, \ldots, X_n] = X_n \mathbb{E}[Z_{n+1}] = X_n,$$

since $\mathbb{E}[Z_{n+1}] = \frac{1}{2}(1.2) + \frac{1}{2}(0.8) = 1$. Therefore $X_n \geq 0$ and the martingale convergence theorem implies that $X_n \to X_\infty$ almost surely for some finite random variable X_∞.

To identify the limit, write $U_n = \sum_{i=1}^{n} \mathbb{1}_{\{Z_i = 1.2\}}$ and $D_n = n - U_n$. Then

$$X_n = a\, 1.2^{U_n} 0.8^{D_n} = a \exp\big(U_n \log(1.2) + D_n \log(0.8)\big).$$

Hence

$$\frac{1}{n} \log\left(\frac{X_n}{a}\right) = \frac{U_n}{n} \log(1.2) + \frac{D_n}{n} \log(0.8) \longrightarrow \frac{1}{2} \log(1.2) + \frac{1}{2} \log(0.8) = \frac{1}{2} \log(0.96) < 0.$$

Therefore $\log(X_n/a) \to -\infty$, so $X_n \to 0$ almost surely. In particular, $X_\infty = 0$ almost surely.

An intuitive way to read this is: in the long run, the number of upward and downward moves is approximately the same, and one up with one down multiplies the price by $1.2 \times 0.8 = 0.96 < 1$. Repeated pairs drive the price downward.

This does not contradict the martingale property. We still have $\mathbb{E}[X_n] = a$ for every n. The mean is preserved by increasingly rare trajectories on which X_n becomes very large, while with overwhelming probability the process drifts toward 0.

Appendix: Lecture and Tutorial Index