Consider a continuous time stochastic process \((X_t, t \geq 0)\) taking on values in the finite state space \(S = \{0, 1, 2, \ldots, N\}\).

Recall that in discrete time, given a transition matrix \(P = p(i, j)\) with \(i, j \in S\), the \(n\)-step transition matrix is simply \(P^n = p^n(i, j)\), and that the following relationship holds:

\[
P^n P^m = P^{n+m}.
\]

Where \(p^n(i, j)\) is the probability that the process moves from state \(i\) to state \(j\) in \(n\) transitions. That is,

\[
\begin{align*}
p^n(i, j) &= \mathbb{P}(X_n = j), \\
p^n(i, j) &= \mathbb{P}(X_n = j|X_0 = i), \\
p^n(i, j) &= \mathbb{P}(X_{m+n} = j|X_m = i).
\end{align*}
\]

Now moving to continuous time, we say that the process \((X_t, t \geq 0)\) is a continuous time Markov chain if the following properties hold for all \(i, j \in S, t, s \geq 0:\)

- \(P_t(i, j) \geq 0,\)
- \(\sum_{j=0}^{N} P_t(i, j) = 1,\)
- \(\mathbb{P}(X_{t+s} = j|X_0 = i) = \sum_{k=0}^{N} \mathbb{P}(X_{t+s} = j|X_s = k)\mathbb{P}(X_s = k|X_0 = i),\) or
- \(\mathbb{P}(X_{t+s} = j|X_0 = i) = \sum_{k=0}^{N} P_t(i, k)P_t(k, j).\)

This last property can be written in matrix form as

\[
P_{s+t} = P_s P_t.
\]

This is known as the Chapman-Kolmogorov equation (the semi-group property).

**Example 21.1 (Poissonize a Discrete Time Markov Chain)** Consider a poisson process \((N_t, t \geq 0)\) with rate \(\lambda\), and a jump chain with transition matrix \(\hat{P}\), which
makes jumps at times \((N_t, t \geq 0)\). Let \(X_t = Y_{N_t}\) (\(Y_{N_t}\) is the value of the jump chain at time \(N_t\)), where \(X_t\) takes values in the discrete state space \(S\). Assume \((Y_0, Y_1, \ldots)\) and \(N_t\) are independent. Find \(P_t\) for the process \(X_t\).

By definition
\[
P_t(i, j) = \mathbb{P}(X_t = j | X_0 = i).
\]
Condition on \(N_t\) to give,
\[
P_t(i, j) = \sum_{n=0}^{\infty} \mathbb{P}(N_t = n) \hat{P}^n(i, j)
= \sum_{n=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} \hat{P}^n(i, j)
= e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n(i, j)}{n!}
= e^{-\lambda t} e^{-\lambda \hat{P}}.
\]
Dropping indices then gives,
\[
P_t = e^{-\lambda t(P-I)}.
\]
\(P_t\) is a paradigm example of a continuous time semi-group of transition matrices. It is easy to check that \(P_t P_s = P_{t+s}\), since \(e^{A+B} = e^A e^B\).

### 21.1 Generalization

Let \(A\) be a matrix of transition rates, where \(a_{ij} = A(i, j)\) is the rate of transitions from state \(i\) to state \(j\) \((i \neq j)\), \(a_{ij} \geq 0\), and \(a_i = A(i, i)\) is the total rate of transitions out of state \(i\). Hence, \(a_i = -\sum_{j=0, j \neq i}^{N} a_{ij}\) and each row of \(A\) sums to 0. Also, since \(P_t\) is continuous and differentiable we have
\[
a_{ij} = \lim_{h \to 0} \frac{P_h(i, j)}{h},
a_i = \lim_{h \to 0} \frac{1 - P_h(i, i)}{h}.
\]
In words, \(a_{ij}\) is the probability per unit time of transitioning from state \(i\) to state \(j\) in a small time increment \((h)\).

The limit relations above can be expressed in matrix form,
\[
\frac{d}{dt} P_t = \lim_{h \to 0^+} \frac{P_{t+h} - P_t}{h} = \lim_{h \to 0^+} \frac{P_tP_h - P_t}{h} = \lim_{h \to 0^+} \frac{P_t[P_h - I]}{h}.
\]
hence
\[ \frac{d}{dt} P_t = P_t \lim_{h \to 0^+} \frac{P_{t+h} - I}{h} = P_t \frac{d}{dt} P_0 = P_t A = AP_t, \]
where \( A = \frac{d}{dt} P_0 \). Given the initial condition \( P_0 = I \), the solution to the equation
\[ \frac{d}{dt} P_t = P_t A = AP_t \]
is \( P_t = e^{At} \). \( \frac{d}{dt} P_t = P_t A \) is known as the Kolmogorov forward equations and \( \frac{d}{dt} P_t = AP_t \) is known as the Kolmogorov backward equations. We claim that \( (P_t, t \geq 0) \) is a collection of transition matrices of some continuous time markov chain.

In the example given above we have,
\[ A = \begin{pmatrix} \lambda & \mu \\ -\lambda & -\mu \end{pmatrix}, \]
\[ A(i,i) = \lambda(\hat{P}(i,i) - 1) \quad \text{and} \quad A(i,j) = \lambda\hat{P}(i,j). \]

**Example 21.2** Consider the two state markov chain with states \( \{0, 1\} \) and
\[ A = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}. \]
The sojourn times in state 0 and 1 are independent and exponentially distributed with parameters \( \lambda \) and \( \mu \), respectively. Find \( P_t \) for this chain.

First note that
\[ \frac{d}{dt} P_t = AP_t = \begin{pmatrix} -\lambda P_t(0,0) + \lambda P_t(1,0) & -\lambda P_t(0,1) + \lambda P_t(1,1) \\ \mu P_t(0,0) - \mu P_t(1,0) & \mu P_t(0,1) - \mu P_t(1,1) \end{pmatrix}. \]
Then
\[ \frac{d}{dt}(P_t(0,0) - P_t(1,0)) = -(\lambda + \mu)(P_t(0,0) - P_t(1,0)). \]
This equation has the form \( f'(t) = cf(t) \), and a solution is
\[ P_t(0,0) - P_t(1,0) = ce^{-(\lambda+\mu)t}. \]
Since \( c \) evaluated at \( t = 0 \) is one (i.e., \( P_0(0,0) - P_0(1,0) = 1 \)), we have
\[ P_t(0,0) - P_t(1,0) = e^{-(\lambda+\mu)t}, \]
which implies
\[ \frac{d}{dt} P_t(0,0) = -\lambda e^{-(\lambda+\mu)t}. \]
Then

\[
P_t(0, 0) = P_0(0, 0) + \int_0^t -\lambda e^{-(\lambda+\mu)s} ds
\]

\[
P_t(0, 0) = 1 + \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)t} - \frac{\lambda}{\lambda + \mu}
\]

\[
P_t(0, 0) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)t}.
\]

Likewise

\[
P_t(1, 0) = \frac{\mu}{\lambda + \mu} - \frac{\mu}{\lambda + \mu} e^{-(\lambda+\mu)t}
\]

\[
P_t(0, 1) = \frac{\mu}{\lambda + \mu} - \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)t}
\]

\[
P_t(1, 1) = \frac{\lambda}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} e^{-(\lambda+\mu)t}.
\]

Notice that as \( t \to \infty \)

\[
P_t(0, 0) = P_t(1, 0) = \frac{\mu}{\lambda + \mu}
\]

\[
P_t(0, 1) = P_t(1, 1) = \frac{\lambda}{\lambda + \mu}.
\]

These limits are independent of the initial state. The stationary distribution of the chain is

\[
\pi_0 = \frac{\mu}{\lambda + \mu}
\]

\[
\pi_1 = \frac{\lambda}{\lambda + \mu}.
\]

### 21.2 Jump Hold Description

Let \((X_t, t \geq 0)\) be a markov chain with transition rate matrix \(A\), such that \(A(i, j) \geq 0\) for \(i \neq j\) and \(A(i, i) = -\sum_{i \neq j} A(i, j)\). Let \(H_i\) be the holding time in state \(i\) (i.e., the process "holds" in state \(i\) for an amount of time \(H_i\) and then jumps to another state). The markov property implies

\[
P_i(H_i > s + t | H_i > s) = P_i(H_i > t).
\]

We also recognize this as the memoryless property of the exponential distribution (fact: the only continuous distribution with the memoryless property is the exponential).
By the idea of *competing risks* we will see that \( \mathbb{P}_i(H_i > t) = e^{A(i,i)t} \). Imagine that each state \( j \neq i \) has a random variable \( \xi_j \sim \exp(A(i,j)) \). Think of a bell that rings in state \( j \) at time \( \xi_j \), and from state \( i \) the process moves to the state where the bell rings first. Construct \( H_i = \min_{i \neq j} \xi_j \), then

\[
\mathbb{P}_i(H_i > t) = \mathbb{P}\left( \bigcap_{i \neq j} \xi_j > t \right) = \prod_{i \neq j} e^{-A(i,j)t} = e^{-(\sum_{i \neq j} A(i,j))t} = e^{A(i,i)t},
\]

where \( A(i,i) < 0 \). What is the probability is that the first transition out of state \( i \) is to state \( k \)? This is given by

\[
\mathbb{P}(\min_{i \neq j} \xi_j = \xi_k) = \int_0^\infty \mathbb{P}(\min \in dt, \ \min = \xi_k)dt = \int_0^\infty \mathbb{P}(\xi_k \in dt, \text{all the others are } > t)dt = \int_0^\infty A(i,k)e^{-A(i,k)t}e^{-(\sum_{i \neq j} A(i,j))t}dt = \int_0^\infty A(i,k)e^{-(\sum_{i \neq j} A(i,j))t}dt = \frac{A(i,k)}{\sum_{i \neq j} A(i,j)} = \frac{A(i,k)}{-A(i,i)}.
\]

Observe from this calculation that the following mechanisms are equivalent:

1. (a) competing random variables \( \xi_j \sim \exp(A(i,j)) \) for state \( j \neq i \)
   (b) \( H_i = \min_{j \neq i} \xi_j \)
   (c) jump to the state \( j \) which attains the minimum

2. (a) \( H_i \sim \exp(-A(i,i) = \sum_{i \neq j} A(i,j)) \)
   (b) at time \( H_i \) jump to \( k \) with probability \( \frac{A(i,k)}{A(i,i)} \).

**Example 21.3 (Hold Rates for a Poissonized Chain)** Consider a general transition matrix \( \hat{P} \), where some of the diagonal entries maybe greater than zero. So we have

\[
P_t = e^{-\lambda(t)} \\
A = \lambda(\hat{P} - I) \\
A(i,i) = \lambda(\hat{P}(i,i) - 1) = -\lambda(1 - \hat{P}(i,i)) \\
A(i,j) = \lambda(\hat{P}(i,j)).
\]
Hence \( A(i, i) = \lambda(1 - \hat{P}(i, i)) = \lambda \sum_{i \neq j} \hat{P}(i, j) = \sum_{i \neq j} A(i, j) \) is the total rate of transitions out of state \( i \), and \( A(i, j) = \lambda(\hat{P}(i, j)) \) is the rate of \( i \) to \( j \) transitions. Let \( H_i = S_1 + S_2 + \cdots + S_G \), where \( S_1, S_2, \ldots \sim \exp(\lambda) \) are independent. Define \( G \) to be the number of steps of the \( \hat{P} \) chain until the process leaves state \( i \). Then

\[
P(G = n) = \hat{P}(i, i)^{n-1}(1 - \hat{P}(i, i)).
\]

So \( G \) has a geometric distribution with parameter \( 1 - \hat{P}(i, i) \) and \( H_i \) has an exponential distribution with parameter \( \lambda(1 - \hat{P}(i, i)) \). It is easy to check the expectations, since \( \mathbb{E}[S_i] = 1/\lambda \), \( \mathbb{E}[G] = 1/(1 - \hat{P}(i, i)) \), and \( \mathbb{E}[H_i] = 1/\lambda(1 - \hat{P}(i, i)) \). Then by Wald’s identity

\[
\mathbb{E}[H_i] = \mathbb{E}[S_i]\mathbb{E}[G] = \frac{1}{\lambda} \left( \frac{1}{1 - \hat{P}(i, i)} \right) = \frac{1}{\lambda(1 - \hat{P}(i, i))}.
\]

### 21.3 Theorem

**Theorem 21.4 (for finite state irreducible chain)** There is a unique probability distribution \( \pi \) so that

1. \( \lim_{t \to \infty} P_t(i, j) = \pi_j \) for all \( i \).
2. \( \pi \) solves the ”balance equations” \( \pi A = 0 \) (equivalently, \( \pi P_t = \pi \) for all \( t \)).
3. \( \pi_j \) is the long-run proportion of time spent in state \( j \).
4. the expected return time is

\[
\mathbb{E}_j[T_j] = \frac{1}{q_j \pi_j}.
\]

**Proof:** [for part (2)] Assume (1) is true.

\[
P_{s+t}(i, j) = \sum_k P_s(i, k)P_t(k, j)
\]

Letting \( s \to \infty \) we get

\[
\pi_j = \sum_k \pi_k P_t(k, j)
\]

Therefore, \( \pi = \pi P_t \) for all \( t \). Differentiating with respect to \( t \) then gives

\[
0 = \pi A.
\]
Look at the \( j \)th element of \( \pi A \) (which we get by multiplying \( \pi \) by the \( j \)th column of \( A \)). We see that

\[
\sum_{i \neq j} \pi_i q_{ij} - \pi_j q_j = 0
\]

which gives

\[
\sum_{i \neq j} \pi_i q_{ij} = \pi_j q_j.
\]

This is the reason these equations are referred to as balance equations. Computations can often be simplified if there is also detailed balance, which simply means that for all pairs \( i, j \ (i \neq j) \) we have

\[
\pi_i q_{ij} = \pi_j q_{ji}.
\]

This is the condition of reversibility of the markov chain.

**Example 21.5 (Birth-Death Process)** Consider a process that, from state \( i \), can only move to state \( i + 1 \) or state \( i - 1 \). That is

\[
q_{i,i+1} = \lambda_i, \quad q_{i,i-1} = \mu_i, \quad q_{i,j} = 0, \forall j \notin \{i + 1, i - 1\}.
\]

So \( \lambda_i \) is the birth rate and \( \mu_i \) is the death rate. We try to find \( \pi \) so that

\[
\pi_0 q_{0,1} = \pi_1 q_{1,0}
\]

\[
\pi_1 q_{1,2} = \pi_2 q_{2,1}
\]

\[
\vdots
\]

\[
\pi_{i-1} q_{i-1,i} = \pi_i q_{i,i-1}.
\]

Then

\[
\pi_0 q_{0,1} = \pi_1 q_{1,0}
\]

\[
\pi_0 \lambda_0 = \pi_1 \mu_1
\]

\[
\pi_1 = \frac{\lambda_0}{\mu_1} \pi_0.
\]

And

\[
\pi_1 q_{1,2} = \pi_2 q_{2,1}
\]

\[
\pi_1 \lambda_1 = \pi_2 \mu_2
\]

\[
\pi_2 = \frac{\lambda_1}{\mu_2} \pi_1 = \frac{\lambda_1 \lambda_0}{\mu_1 \mu_2} \pi_0.
\]
Hence

\[ \pi_k = \frac{\lambda_0 \lambda_1 \cdots \lambda_{k-1}}{\mu_1 \mu_2 \cdots \mu_k} \pi_0, \]

and

\[ \pi_0 = \frac{1}{1 + \sum_{k=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{k-1}}{\mu_1 \mu_2 \cdots \mu_k}}. \]