MTH 665

Lectures 19 - 23

Yevgeniy Kovchegov
Oregon State University
Topics:

- Reversible Markov chains: applications
- Markov Chain Monte Carlo (MCMC)
- Mixing times
- Coupling method
- Continuous-time Markov chains
- Poisson process
Metropolis-Hastings algorithm.

**Goal:** simulating an $S$-valued random variable distributed according to a given probability distribution $\pi(z)$, given a complex nature of large discrete space $S$.

**MCMC:** generating a Markov chain $\{X_t\}$ over $S$, with distribution $\mu_t(z) = P(X_t = z)$ converging rapidly to its unique stationary distribution, $\pi(z)$.

**Metropolis-Hastings algorithm:** Consider a connected neighborhood network with points in $S$. Suppose we know the ratios of $\frac{\pi(z')}{\pi(z)}$ for any two neighbor points $z$ and $z'$ on the network.

Let for $z$ and $z'$ connected by an edge of the network, the transition probability be set to

$$p(z, z') = \frac{1}{M} \min \left\{ 1, \frac{\pi(z')}{\pi(z)} \right\} \quad \text{for } M \text{ large enough.}$$
Metropolis-Hastings algorithm.

Consider a connected neighborhood network with points in $S$.

Suppose we know the ratios of $\frac{\pi(z')}{\pi(z)}$ for any two neighbor points $z$ and $z'$ on the network.

Let for $z$ and $z'$ connected by an edge of the network, the transition probability be set to

$$p(z, z') = \frac{1}{M} \min \left\{ 1, \frac{\pi(z')}{\pi(z)} \right\} \text{ for } M \text{ large enough.}$$

Specifically, $M$ can be any number greater than the maximal degree in the neighborhood network.

Let $p(z, z)$ absorb the rest of the probabilities, i.e.

$$p(z, z) = 1 - \sum_{z': z \sim z'} p(z, z')$$
**Knapsack problem.** The **knapsack problem** is a problem in combinatorial optimization: Given a set of items, each with a mass and a value, determine the number of each item to include in a collection so that the total weight is less than or equal to a given limit and the total value is as large as possible. Knapsack problem is NP complete.

Source: Wikipedia.org
Knapsack problem. Given \( m \) items of various weights \( w_j \) and value \( v_j \), and a knapsack with a weight limit \( R \). Assuming the volume and shape do not matter, find the most valuable subset of items that can be carried in the knapsack.

Mathematically: we need \( z = (z_1, \ldots, z_m) \) in

\[
S = \{ z \in \{0, 1\}^m : \sum_{j=1}^{m} w_j z_j \leq R \}
\]

maximizing \( U(z) = \sum_{j=1}^{m} v_j z_j \).

Source: Wikipedia.org
Knapsack problem. Find \( z = (z_1, \ldots, z_m) \) in

\[
S = \{ z \in \{0, 1\}^m : \sum_{j=1}^{m} w_j z_j \leq R \} \text{ maximizing } U(z) = \sum_{j=1}^{m} v_j z_j.
\]

• MCMC approach: Assign weights \( \pi(z) = \frac{1}{Z_\beta} \exp \{ \beta U(z) \} \) to each \( z \in S \) with \( \beta = \frac{1}{T} \), where

\[
Z_\beta = \sum_{z \in S} \exp \{ \beta U(z) \}
\]

is called partition function. Next, for each \( z \in S \) consider a clique \( C_z \) of neighbor points in \( S \). Consider a Markov chain over \( S \) that jumps from \( z \) to a neighbor \( z' \in C_z \) with probability

\[
p(z, z') = \frac{1}{m} \min \left\{ 1, \frac{\pi(z')}{\pi(z)} \right\}.
\]
Knapsack problem. Assign weights $\pi(z) = \frac{1}{Z_\beta} \exp \{\beta \ U(z)\}$ to each $z \in S$ with $\beta = \frac{1}{T}$, where

$$Z_\beta = \sum_{z \in S} \exp \{\beta \ U(z)\}$$

is called partition function. Next, for each $z \in S$ consider a clique $C_z$ of neighbor points in $S$. Consider a Markov chain over $S$ that jumps from $z$ to a neighbor $z' \in C_z$ with probability

$$p(z, z') = \frac{1}{m} \ \min \left\{ 1, \frac{\pi(z')}{\pi(z)} \right\}.$$ 

Observe that

$$\frac{\pi(z')}{\pi(z)} = \exp \{\beta \ (U(z') - U(z))\} = \exp \{\beta \ (v \cdot (z' - z))\},$$

where $v = (v_1, \ldots, v_m)$ is the values vector.
Knapsack and other optimization problems.

• Issues:

(i) Running time?
Analyzing mixing time is challenging in MCMC for real-life optimization problems such as knapsack problem. With few exceptions – no firm foundation exists, and no performance guaranteed.

(ii) Optimal $T$?
$T$ is usually chosen using empirical observations, trial and error, or certain heuristic.
Often, simulated annealing approach is used.
Simulated annealing.

Usually, we let \( \pi(z) = \frac{1}{Z_\beta} \exp \{ \beta \ U(z) \} \) to each \( z \in S \) with \( \beta = \frac{1}{T} \), and \( p(z, z') = \frac{1}{M} \ \min \left\{ 1, \frac{\pi(z')}{\pi(z)} \right\} \).

- **Idea:** What if we let temperature \( T \) change with time \( t \), i.e. \( T = T(t) \)? When \( T \) is large, the Markov chain is more diffusive; as \( T \) gets smaller, the value \( X_t \) stabilizes around the maxima.


Name comes from *annealing in metallurgy*, a technique involving heating and controlled cooling.
Gibbs Sampling: Ising Model.

Every vertex $v$ of $G = (V, E)$ is assigned a spin $\sigma(v) \in \{-1, +1\}$. The probability of a configuration $\sigma \in \{-1, +1\}^V$ is

$$\pi(\sigma) = \frac{e^{-\beta \mathcal{H}(\sigma)}}{Z(\beta)}, \quad \text{where} \quad \beta = \frac{1}{T}$$
Gibbs Sampling: Ising Model.

Every vertex $v$ of $G = (V, E)$ is assigned a spin $\sigma(v) \in \{-1, +1\}$. The probability of a configuration $\sigma \in \{-1, +1\}^V$ is

$$\pi(\sigma) = \frac{e^{-\beta H(\sigma)}}{Z(\beta)}, \quad \text{where} \quad \beta = \frac{1}{T}$$
**Gibbs Sampling: Ising Model.**

∀σ ∈ \{-1, +1\}^V, the Hamiltonian (energy function)

\[
H(\sigma) = -\frac{1}{2} \sum_{u,v: u \sim v} \sigma(u)\sigma(v) = - \sum_{\text{edges } e=[u,v]} \sigma(u)\sigma(v)
\]

and probability of a configuration σ ∈ \{-1, +1\}^V is

\[
\pi(\sigma) = \frac{e^{-\beta H(\sigma)}}{Z(\beta)}, \quad \text{where } \beta = \frac{1}{T}
\]

\[
Z(\beta) = \sum_{\sigma \in \{-1,+1\}^V} e^{-\beta H(\sigma)} - \text{normalizing factor.}
\]

The local Hamiltonian \( H_{local}(\sigma,v) = - \sum_{u: u \sim v} \sigma(u)\sigma(v) \).

Conditional probability for \( \sigma(v) \) is expressed via \( H_{local}(\sigma,v) \):

\[
H(\sigma) = H_{local}(\sigma,v) - \sum_{e=[u_1,u_2]: u_1,u_2 \neq v} \sigma(u_1)\sigma(u_2)
\]
Gibbs Sampling: Ising Model via Glauber dynamics.

Conditional probability for $\sigma(v)$ is expressed via $H_{local}(\sigma, v)$:

$$H(\sigma) = H_{local}(\sigma, v) - \sum_{e = [u_1, u_2]: u_1, u_2 \neq v} \sigma(u_1)\sigma(u_2)$$
Gibbs Sampling: Ising Model via Glauber dynamics.

Randomly pick $v \in V$ with probability $\frac{\text{deg}(v)}{2|E|}$. Next, erase the spin $\sigma(v)$, and replace it with $+1$ or $-1$ with probabilities

$$P(\sigma \to \sigma_+) = \frac{e^{-\beta H(\sigma_+)}}{e^{-\beta H(\sigma_-)} + e^{-\beta H(\sigma_+)}} = \frac{e^{-\beta H_{\text{local}}(\sigma_+,v)}}{e^{-\beta H_{\text{local}}(\sigma_-,v)} + e^{-\beta H_{\text{local}}(\sigma_+,v)}} = \frac{e^{-2\beta \sigma(v)}}{e^{-2\beta \sigma(v)} + e^{2\beta \sigma(v)}}$$

and

$$P(\sigma \to \sigma_-) = \frac{e^{2\beta \sigma(v)}}{e^{-2\beta \sigma(v)} + e^{2\beta \sigma(v)}}.$$
Gibbs Sampling: Ising Model via Glauber dynamics.
Randomly pick $v \in V$ with probability $\frac{\text{deg}(v)}{2|E|}$. Next, erase the spin $\sigma(v)$, and replace it with $+1$ or $-1$ with probabilities

$$P(\sigma \rightarrow \sigma_+) = \frac{e^{-\beta \mathcal{H}(\sigma_+)} - \beta \mathcal{H}_{\text{local}}(\sigma_+,v)}{e^{-\beta \mathcal{H}(\sigma_-)} + e^{-\beta \mathcal{H}(\sigma_+)} - \beta \mathcal{H}_{\text{local}}(\sigma_-,v) + e^{-\beta \mathcal{H}(\sigma_+)}} = \frac{e^{-\beta \mathcal{H}(\sigma_+)} - \beta \mathcal{H}_{\text{local}}(\sigma_+,v)}{e^{-\beta \mathcal{H}(\sigma_-)} + e^{-\beta \mathcal{H}(\sigma_+)} - \beta \mathcal{H}_{\text{local}}(\sigma_-,v) + e^{-\beta \mathcal{H}(\sigma_+)}}$$

and

$$P(\sigma \rightarrow \sigma_-) = \frac{e^{2\beta \sigma(v)}}{e^{-2\beta \sigma(v)} + e^{2\beta \sigma(v)}}.$$  

Here, $\sigma_+,v = \sigma_+$ and $\sigma_-,v = \sigma_-$ are given by

$$\sigma_+(u) = \begin{cases} \sigma(u) & \text{if } u \neq v, \\ +1 & \text{if } u = v \end{cases}$$

and

$$\sigma_-(u) = \begin{cases} \sigma(u) & \text{if } u \neq v, \\ -1 & \text{if } u = v \end{cases}$$

So, the transition probabilities are

$$p(\sigma, \sigma_+,v) = \frac{\text{deg}(v)}{2|E|} \frac{e^{2\beta \sigma(v)}}{e^{-2\beta \sigma(v)} + e^{2\beta \sigma(v)}}$$

and

$$p(\sigma, \sigma_-,v) = \frac{\text{deg}(v)}{2|E|} \frac{e^{2\beta \sigma(v)}}{e^{-2\beta \sigma(v)} + e^{2\beta \sigma(v)}}.$$
Glauber dynamics: Rapid mixing.

Glauber dynamics - a random walk on state space $S$ (here $\{-1, +1\}^V$) s.t. needed $\pi$ is stationary w.r.t. Glauber dynamics.

In high temperatures (i.e. $\beta = \frac{1}{T}$ small enough) it takes $O(n \log n)$ iterations to get “$\varepsilon$-close” to $\pi$. Here $|V| = n$.

Need: $\max_{v \in V} \text{deg}(v) \cdot \tanh(\beta) < 1$

Thus the Glauber dynamics is a fast way to generate $\pi$. It is an important example of Gibbs sampling.
Close enough distribution and mixing time.

What is “$\varepsilon$-close” to $\pi$? Start with $\sigma_0$:

\[
\begin{array}{cccccccc}
- & +1 & - & +1 & - & +1 & - & +1 \\
- & +1 & - & +1 & - & +1 & - & +1 \\
- & +1 & - & +1 & - & -1 & - & -1 \\
- & -1 & - & -1 & - & -1 & - & -1 \\
\end{array}
\]

If $P_t(\sigma)$ is the probability distribution after $t$ iterations, the total variation distance

\[
\|P_t - \pi\|_{TV} = \frac{1}{2} \sum_{\sigma \in \{-1,+1\}^V} |P_t(\sigma) - \pi(\sigma)| \leq \varepsilon .
\]
Close enough distribution and mixing time.

**Total variation distance:**
\[ \|\mu - \nu\|_{TV} := \frac{1}{2} \sum_{x \in S} |\mu(x) - \nu(x)| = \sup_{A \subset S} |\mu(A) - \nu(A)|. \]

**Mixing time:** let \( \mu_t = \mu_0 P^t \), then
\[ t_{mix}(\varepsilon) := \inf \{ t : \|\mu_t - \pi\|_{TV} \leq \varepsilon, \text{ all } \mu_0 \}. \]

In high temperature, \( t_{mix}(\varepsilon) = O(n \log n) \).
**Coupling Method.**

$S$ - sample space

$\{p(i, j)\}_{i, j \in S}$ - transition probabilities

Construct process $\begin{pmatrix} X_t \\ Y_t \end{pmatrix}$ on $S \times S$ such that

$X_t$ is a $\{p(i, j)\}$-Markov chain

$Y_t$ is a $\{p(i, j)\}$-Markov chain

Once $X_t = Y_t$, let $X_{t+1} = Y_{t+1}$, $X_{t+2} = Y_{t+2}, \ldots$
**Coupling Method.**

**Coupling time:** \( \tau = \min\{t : X_t = Y_t\} \)

**Successful coupling:** \( \text{Prob}(\tau < \infty) = 1 \)
Mixing times via coupling. For any given $x \in S$, let $X_0 = x$, and $Y_0$ be distributed with probabilities $P(Y_0 = y) = \pi(y)$ $\forall y \in S$. Then, for all $n \in \mathbb{N}$, $P(Y_n = y) = \pi(y)$ $\forall y \in S$. Hence,

$$\sum_{y \in S} |p_t(x, y) - \pi(y)| = \sum_{y \in S} |P(X_t = y) - P(Y_t = y)|$$

$$= \sum_{y \in S} |P(X_t = y, X_t \neq Y_t) - P(Y_t = y | X_t \neq Y_t)| \leq 2P(\tau > t)$$

Thus, by convexity of $\| \cdot \|_{TV}$, if $X_0$ is $\mu_0$-distributed, and $\mu_t = \mu_0 P_t$, then

$$\|\mu_t - \pi\|_{TV} \leq \frac{2}{t} \max_{i,j \in S} E_{i,j}[\tau] \leq \varepsilon \quad \text{whenever} \quad t \geq \frac{2}{\varepsilon} \max_{i,j \in S} E_{i,j}[\tau].$$

So, $O(t_{mix}) \leq O(\tau)$ as

$$t_{mix}(\varepsilon) = \inf \{ t : \|P_{X_t} - \pi\|_{TV} \leq \varepsilon \} \leq \frac{\max_{i,j \in S} E_{i,j}[\tau]}{\varepsilon}. $$
Coupon collector.

$n$ types of coupons: $1, 2, \ldots, n$. Collecting coupons: coupon / unit of time, each coupon type is equally likely. Goal: To collect a coupon of each type. Question: How much time will it take?

Here, $\tau_1 = 1$, $E[\tau_2-\tau_1] = \frac{n}{n-1}$, $E[\tau_3-\tau_2] = \frac{n}{n-2}$, $\ldots$, $E[\tau_n-\tau_{n-1}] = n$.

Hence,

$$E[\tau_n] = n \left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right) = n \log n + O(n)$$
Continuous-time Markov chains.

Consider a family of random variables (process) \( \{X(t)\} \) with \( t \in [0, \infty) \), taking values in a discrete state space \( S \). W.l.o.g. we enumerate the state space by letting \( S = \{0, 1, 2, \ldots\} \).

Process \( X(t) \) is a continuous-time Markov chain if it satisfies the following continuous Markov property:

\[
P(X(t) = j \mid X(t_n) = i_n, X(t_{n-1}) = i_{n-1}, \ldots, X(t_1) = i_1) = P(X(t) = j \mid X(t_n) = i_n)
\]

for any sequence

\[
0 \leq t_1 < t_2 < \ldots < t_n < t.
\]

Continuous-time Markov chain \( X(t) \) is said to be time homogeneous if there are probabilities \( p_t(i, j) \) with \( t \in [0, \infty) \) and \( i, j \in S \) such that

\[
P(X(t) = j \mid X(s) = i) = p_{t-s}(i, j)
\]

for all \( 0 \leq s < t, \ i, j \in S \).
Continuous-time Markov chains.

Consider a time homogeneous continuous-time Markov chain $X(t)$.

$$P(X(t) = j \mid X(s) = i) = p_{t-s}(i, j) \quad \text{for all} \ 0 \leq s < t, \ i, j \in S.$$  

For $t \in [0, \infty)$, let

$$P_t = \left( p_t(i, j) \right)_{i,j \in S}$$

Observe that $P_0 = I$, and $P_t$ is stochastic, i.e., non-negative with rows adding to one.

Chapman-Kolmogorov equation.

$$P_{s+t} = P_s P_t, \quad \forall s, t \in [0, \infty)$$

i.e., $\{P_t\}_{t \geq 0}$ form a semigroup.
Continuous-time Markov chains. \( P_0 = I \), and

\[
P_{s+t} = P_s P_t, \quad \forall s, t \in [0, \infty)
\]

i.e., \( \{P_t\}_{t \geq 0} \) form a semigroup.

Generator:

\[
G = \lim_{h \downarrow 0} \frac{P_h - I}{h}
\]

Forward equations:

\[
\frac{d}{dt} P_t = P_t G, \quad P_0 = I
\]

Backward equations:

\[
\frac{d}{dt} P_t = G P_t, \quad P_0 = I
\]

Representation:

\[
P_t = \exp\{tG\} = \sum_{k=0}^{\infty} \frac{1}{k!} t^k G^k, \quad t \in [0, \infty)
\]
**Poisson process.**

Let $X_1, X_2, \ldots$ be independent exponential random variables with parameter $\lambda > 0$.

- **Arrival times:** Let $T_0 = 0$ and
  \[
  T_n = \sum_{k=1}^{n} X_k \quad \text{for} \quad n = 1, 2, \ldots.
  \]

- **Interarrival times:** $X_n = T_n - T_{n-1}$

- **Poisson process** with intensity $\lambda$ is defined as
  \[
  N(t) = \max\{n \geq 0 : T_n \leq t\} \quad (t \geq 0).
  \]

Here $T_n$ represent the arrival times and $N(t)$ counts the number of arrivals between 0 and $t$. 
Poisson process.

- Poisson process with intensity $\lambda$ is defined as
  \[
  N(t) = \max\{n \geq 0 : T_n \leq t\} \quad (t \geq 0).
  \]

Here $T_n$ represent the arrival times and $N(t)$ counts the number of arrivals between 0 and $t$.

- The increment $N(t_0 + L) - N(t_0)$ counts the number of arrivals between $t_0$ and $t_0 + L$.

- Because of memorylessness property of exponential random variables $X_j$, the increment $N(t_0 + L) - N(t_0)$ is distributed as $N(L)$.

- $N(t_0 + L) - N(t_0)$ is a Poisson random variable with parameter $\lambda L$:
  \[
P\left(N(t_0 + L) - N(t_0) = k\right) = e^{-\lambda L} \frac{(\lambda L)^k}{k!} \quad (k = 0, 1, \ldots).
  \]
Poisson process.

\[ P\left(N(t_0+L)-N(t_0) = k\right) = e^{-\lambda L} \frac{(\lambda L)^k}{k!} \quad (k = 0, 1, \ldots) \]

**Proof:** Recall \( N(t_0+L)-N(t_0) \) is distributed as \( N(L) \).

Now, since \( \{T_k \leq L\} = \{T_k \leq L < T_{k+1}\} \cup \{T_{k+1} \leq L\} \),

\[ P(N(L) = k) = P(T_k \leq L < T_{k+1}) = P(T_k \leq L) - P(T_{k+1} \leq L) \]

Since \( T_k = X_1 + \ldots + X_k \) is a gamma random variable with parameters \((\lambda, k)\),

\[ P(N(L) = k) = \int_0^L \frac{\lambda^k x^{k-1}}{(k-1)!} e^{-\lambda x} \, dx - \int_0^L \frac{\lambda^{k+1} x^k}{k!} e^{-\lambda x} \, dx \]
Poisson process.

\[ P\left( N(t_0+L) - N(t_0) = k \right) = e^{-\lambda L} \frac{(\lambda L)^k}{k!} \quad (k = 0, 1, \ldots). \]

Proof (continued):

\[ P(N(L) = k) = \int_0^L \frac{\lambda^k x^{k-1}}{(k-1)!} e^{-\lambda x} \, dx - \int_0^L \frac{\lambda^{k+1} x^k}{k!} e^{-\lambda x} \, dx \]

Integration by parts:

\[ \int_0^L \frac{x^{k-1}}{(k-1)!} e^{-\lambda x} \, dx = e^{-\lambda x} \frac{x^k}{k!} \bigg|_0^L + \int_0^L \frac{x^k}{k!} \lambda e^{-\lambda x} \, dx \]

Plugging in: \[ P(N(L) = k) = e^{-\lambda L} \frac{(\lambda L)^k}{k!} \]
Poisson process.

\[ P\left( N(t_0 + L) - N(t_0) = k \right) = e^{-\lambda L} \frac{(\lambda L)^k}{k!} \quad (k = 0, 1, \ldots). \]

Thus, \( p_t(i, j) = 0 \) if \( j < i \), and

\[ p_t(i, j) = e^{-\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!}, \quad 0 \leq i \leq j < \infty. \]

So, as \( h \to 0+ \), we have

- \( p_h(i, i) = e^{-\lambda h} = 1 - \lambda h + o(h); \)
- \( p_h(i, i + 1) = e^{-\lambda h} \lambda h = \lambda h + o(h); \)
- \( \sum_{k=2}^{\infty} p_h(i, i + k) = o(h). \)
Poisson process.

- \( p_h(i, i) = e^{-\lambda h} = 1 - \lambda h + o(h); \)
- \( p_h(i, i + 1) = e^{-\lambda h} \lambda h = \lambda h + o(h); \)
- \( p_h(i, i + k) = o(h) \) for all \( k \geq 2. \)

Generator:

\[
G = \lim_{h \downarrow 0} \frac{P_h - I}{h} = \begin{pmatrix}
-\lambda & \lambda & 0 & \cdots \\
0 & -\lambda & \lambda & \cdots \\
0 & 0 & -\lambda & \cdots \\
\vdots & \vdots & \ddots & \ddots \\
\end{pmatrix}
\]

\( G = (g_{i,j})_{i,j \geq 0} \) satisfies \( g_{i,j} = \begin{cases} 
-\lambda & \text{if } j = i, \\
\lambda & \text{if } j = i + 1, \\
0 & \text{otherwise}.
\end{cases} \)
Continuous-time Markov chains: generator.

Properties of generators \( G = \lim_{h \downarrow 0} \frac{P_h - I}{h} \)

- Rows of \( G \) add up to one;
- Diagonal elements of \( G \) are nonpositive;
- Off-diagonal elements of \( G \) are nonnegative.

Stationary distribution:

\[ \pi G = 0 \iff \pi P_t = \pi \quad \forall t \in [0, \infty) \]

Ergodicity: Under irreducibility and positive recurrence conditions, there exists a unique stationary distribution \( \pi \), and

\[ \lim_{t \to \infty} p_t(i, j) = \pi(j) \quad \forall i, j \in S. \]
Continuous-time Markov chains.

Example. Consider $S = \{0, 1\}$, and for $\lambda, \mu > 0$, let

$$G = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}.$$ 

Interpretation: Process $X_t$ jumps from state 0 to state 1 after waiting for an interarrival time, distributed as an exponential random variable with parameter $\lambda > 0$. Process $X_t$ jumps from state 1 to state 0 after waiting for an interarrival time, distributed as an exponential random variable with parameter $\mu > 0$. The interarrival times are independent random variables.

We want to find $P_t$. We use backward equations:

$$\frac{d}{dt}P_t = GP_t, \quad P_0 = I$$
Continuous-time Markov chains.

Example (continued). Consider $S = \{0, 1\}$, and for $\lambda, \mu > 0$, let

$$G = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}.$$ 

We want to find $P_t$. We use backward equations:

$$\frac{d}{dt}P_t = GP_t, \quad P_0 = I.$$ 

Denote $\varphi(t) = p_t(0, 1)$ and $\psi(t) = p_t(1, 0)$, then

$$P_t = \begin{pmatrix} 1 - \varphi(t) & \varphi(t) \\ \psi(t) & 1 - \psi(t) \end{pmatrix}$$ 

and

$$\begin{pmatrix} -\varphi'(t) & \varphi'(t) \\ \psi'(t) & -\psi'(t) \end{pmatrix} = \frac{d}{dt}P_t = GP_t = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix} \begin{pmatrix} 1 - \varphi(t) & \varphi(t) \\ \psi(t) & 1 - \psi(t) \end{pmatrix}.$$
Continuous-time Markov chains.

Example (continued).
Denote $\varphi(t) = p_t(0,1)$ and $\psi(t) = p_t(1,0)$, then

$$\begin{pmatrix} -\varphi'(t) & \varphi'(t) \\ \psi'(t) & -\psi'(t) \end{pmatrix} = \frac{d}{dt}P_t = GP_t = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix} \begin{pmatrix} 1 - \varphi(t) & \varphi(t) \\ \psi(t) & 1 - \psi(t) \end{pmatrix}.$$

Hence,

$$\varphi'(t) = \lambda(1 - \varphi(t) - \psi(t))$$
and
$$\psi'(t) = \mu(1 - \varphi(t) - \psi(t))$$

with the initial conditions $\varphi(0) = \psi(0) = 0$ as $P_0 = I$.

First,

$$\frac{d}{dt}(\mu \varphi(t) - \lambda \psi(t)) = 0,$$

implies $\psi(t) = \frac{\mu}{\lambda} \varphi(t)$. Substituting, we have

$$\varphi'(t) = \lambda - (\lambda + \mu)\varphi(t), \quad \varphi(0) = 0.$$
Continuous-time Markov chains.

Example (continued).
Denote $\varphi(t) = p_t(0,1)$ and $\psi(t) = p_t(1,0)$, then

\[\varphi'(t) = \lambda (1 - \varphi(t) - \psi(t)) \quad \text{and} \quad \psi'(t) = \mu (1 - \varphi(t) - \psi(t))\]

with the initial conditions $\varphi(0) = \psi(0) = 0$ as $P_0 = I$.

We deduce $\psi(t) = \frac{\mu}{\lambda} \varphi(t)$. Substituting, we have

\[\varphi'(t) = \lambda - (\lambda + \mu) \varphi(t), \quad \varphi(0) = 0.\]

Solving it, obtain

\[p_t(0, 1) = \varphi(t) = \frac{\lambda}{\lambda + \mu} \left( 1 - e^{-(\lambda + \mu)t} \right)\]

and

\[p_t(1, 0) = \psi(t) = \frac{\mu}{\lambda} \varphi(t) = \frac{\mu}{\lambda + \mu} \left( 1 - e^{-(\lambda + \mu)t} \right).\]
Continuous-time Markov chains.

Example (continued). Consider $S = \{0, 1\}$, and for $\lambda, \mu > 0$, let

$$G = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}.$$  

We find

$$P_t = \begin{pmatrix} \frac{\mu}{\lambda+\mu} \left(1 + \frac{\lambda}{\mu} e^{-(\lambda+\mu)t}\right) & \frac{\lambda}{\lambda+\mu} \left(1 - e^{-(\lambda+\mu)t}\right) \\ \frac{\mu}{\lambda+\mu} \left(1 - e^{-(\lambda+\mu)t}\right) & \frac{\lambda}{\lambda+\mu} \left(1 + \frac{\mu}{\lambda} e^{-(\lambda+\mu)t}\right) \end{pmatrix}.$$  

Notice that

$$\pi = \begin{pmatrix} \frac{\mu}{\lambda+\mu} \\ \frac{\lambda}{\lambda+\mu} \end{pmatrix}$$

is the stationary distribution for the Markov chain. Observe ergodicity:

$$\lim_{t \to \infty} p_t(i, j) = \pi(j) \quad \forall i, j \in S.$$