Continuous Time Markov Chains

Thomas J. Sargent & John Stachurski

Mar 24, 2025

CONTENTS

Authors: Thomas J. Sargent and John Stachurski

These lectures provides a short introduction to continuous time Markov chains. Mathematical ideas are combined with computer code to build intuition and bridge the gap between theory and applications. There are many solved exercises.

The presentation is rigorous but aims toward applications rather than mathematical curiosities (which are plentiful, if one starts to look). Applications are drawn from economics, finance and operations research. I assume readers have some knowledge of discrete time Markov chains. Later lectures, use a small amount of analysis in Banach space.

Code is written in Python and accelerated using JIT compilation via Numba. QuantEcon provides an introduction to these topics.

CONTENTS

MEMORYLESS DISTRIBUTIONS

In addition to what's in Anaconda, this lecture will need the following libraries:

```
!pip install quantecon
```

1.1 Overview

Markov processes are, by definition, forgetful.

In particular, for any Markov processes, the distribution over future outcomes depends only on the current state, rather than the entire history.

In the case of continuous time Markov chains, which jump between discrete states, this requires that the amount of elapsed time since the last jump is not helpful in predicting the timing of the next jump.

In other words, the jump times are "memoryless".

It is remarkable that the only distribution on $\mathbb{R}_+$ with this property is the exponential distribution.

Similarly, the only memoryless distribution on $\mathbb{Z}_+$ is the geometric distribution.

This lecture tries to clarify these ideas.

We will use the following imports:

```
import numpy as np
import matplotlib.pyplot as plt
import quantecon as qe
from numba import njit
from scipy.special import factorial, binom
```

1.2 The Geometric Distribution

Consider betting on a roulette wheel and suppose that red has come up four times in a row.

Since five reds in a row is an unlikely event, many people instinctively feel that black is more likely on the fifth spin — "Surely black will come up this time!"

But rational thought tells us such instincts are wrong: the four previous reds make no difference to the outcome of the next spin.

(Many casinos offer an unlimited supply of free alcoholic beverages in order to discourage this kind of rational analysis.)

A mathematical restatement of this phenomenon is: the geometric distribution is memoryless.

1.2.1 Memorylessness

Let X be a random variable supported on the nonnegative integers $\mathbb{Z}_+$.

We say that X is geometrically distributed if, for some θ satisfying $0 \leq \theta \leq 1$,

$$\mathbb{P}\{X = k\} = (1 - \theta)^k \theta \qquad (k = 0, 1, ...) \tag{1.1}$$

An example can be constructed from the discussion of the roulette wheel above.

Suppose that,

- the outcome of each spin is either red or black,

- spins are labeled by $0, 1, 2, ...,$

- on each spin, black occurs with probability θ and

- outcomes across spins are independent.

Then (1.1) is the probability that the first occurrence of black is at spin k.

(The outcome "black" fails k times and then succeeds.)

Consistent with our discussion in the introduction, the geometric distribution is **memoryless**.

In particular, given any nonnegative integer m, we have

$$\mathbb{P}\{X = m + 1 \mid X > m\} = \theta \tag{1.2}$$

In other words, regardless of how long we have seen only red outcomes, the probability of black on the next spin is the same as the unconditional probability of getting black on the very first spin.

To establish (1.2), we use basic properties of the geometric distribution to obtain.

$$\frac{\mathbb{P}\{X = m + 1 \text{ and } X > m\}}{\mathbb{P}\{X \geq m\}} = \frac{\mathbb{P}\{X = m + 1\}}{\mathbb{P}\{X > m\}} = \frac{(1 - \theta)^{m+1}\theta}{(1 - \theta)^{m+1}} = \theta$$

1.3 The Exponential Distribution

Later, when we construct continuous time Markov chains, we will need to specify the distribution of the holding times, which are the time intervals between jumps.

As discussed above (and again below), the holding time distribution must be memoryless, so that the chain satisfies the Markov property.

While the geometric distribution is memoryless, its discrete support makes it a poor fit for the continuous time case.

Hence we turn to the exponential distribution, which is supported on $\mathbb{R}_+$.

A random variable Y on $\mathbb{R}_+$ is called **exponential with rate** λ, denoted by $Y \sim \text{Exp}(\lambda)$, if

$$\mathbb{P}\{Y > y\} = e^{-\lambda y} \qquad (y \geq 0)$$

1.3.1 From Geometric to Exponential

The exponential distribution can be regarded as the "limit" of the geometric distribution.

To illustrate, let us suppose that

- customers enter a shop at discrete times $t_0, t_1, \ldots$
- these times are evenly spaced, so that $h = t_{i+1} - t_i$ for some $h > 0$ and all $i \in \mathbb{Z}_+$
- at each t_i, either zero or one customers enter (no more because h is small)
- entry at each t_i occurs with probability λh and is independent over i.

The fact that the entry probability is proportional to h is important in what follows.

You can imagine many customers passing by the shop, each entering independently.

If we halve the time interval, then we also halve the probability that a customer enters.

Let

- Y be the time of the first arrival at the shop,
- t be a given positive number and
- $i(h)$ be the largest integer such that $t_{i(h)} \le t$.

Note that, as $h \to 0$, the grid becomes finer and $t_{i(h)} = i(h)h \to t$.

Writing $i(h)$ as i and using the geometric distribution, the probability that the first arrival occurs after t_i is $(1 - \lambda h)^i$.

Hence

$$\mathbb{P}\{Y > t_i\} = (1 - \lambda h)^i = \left(1 - \frac{\lambda i h}{i}\right)^i$$

Using the fact that $e^x = \lim_{i \to \infty} (1 + x/i)^i$ for all x and $ih = t_i \to t$, we obtain, for large i,

$$\mathbb{P}\{Y > t\} \approx e^{-\lambda t}$$

In this sense, the exponential is the limit of the geometric distribution.

1.3.2 Memoryless Property of the Exponential Distribution

The exponential distribution is the only memoryless distribution supported on $\mathbb{R}_+$, as the next theorem attests.

Theorem 1.1 (Characterization of the Exponential Distribution)

If X is a random variable supported on $\mathbb{R}_+$, then there exists a $\lambda > 0$ such that $X \sim \operatorname{Exp}(\lambda)$ if and only if, for all positive s, t,

$$\mathbb{P}\{X > s + t \mid X > s\} = \mathbb{P}\{X > t\} \tag{1.3}$$

Proof. To see that (1.3) holds when X is exponential with rate λ, fix $s, t > 0$ and observe that

$$\frac{\mathbb{P}\{X > s + t \text{ and } X > s\}}{\mathbb{P}\{X > s\}} = \frac{\mathbb{P}\{X > s + t\}}{\mathbb{P}\{X > s\}} = \frac{e^{-\lambda s - \lambda t}}{e^{-\lambda s}} = e^{-\lambda t}$$

To see that the converse holds, let X be a random variable supported on $\mathbb{R}_+$ such that (1.3) holds.

The "exceedance" function $f(s) := \mathbb{P}\{X > s\}$ then has three properties:

1. f is decreasing on $\mathbb{R}_+$,

2. $0 < f(t) < 1$ for all $t > 0$,

3. $f(s + t) = f(s)f(t)$ for all $s, t > 0$.

The first property is common to all exceedance functions, the second is due to the fact that X is supported on all of $\mathbb{R}_+$, and the third is (1.3).

From these three properties we will show that

$$f(t) = f(1)^t \quad \forall\, t \geq 0 \tag{1.4}$$

This is sufficient to prove the claim because then $\lambda := -\ln f(1)$ is a positive real number (by property 2) and, moreover,

$$f(t) = \exp(\ln(f(1))t) = \exp(-\lambda t)$$

To see that (1.4) holds, fix positive integers m, n.

We can use property 3 to obtain both

$$f(m/n) = f(1/n)^m \quad \text{and} \quad f(1) = f(1/n)^n$$

It follows that $f(m/n)^n = f(1/n)^{mn} = f(1)^m$ and, raising to the power of $1/n$, we get (1.4) when $t = m/n$.

The discussion so far confirms that (1.4) holds when t is rational.

So now take any $t \geq 0$ and rational sequences (a_n) and (b_n) converging to t with $a_n \leq t \leq b_n$ for all n.

By property 1 we have $f(b_n) \leq f(t) \leq f(a_n)$ for all n, so

$$f(1)^{b_n} \leq f(t) \leq f(1)^{a_n} \quad \forall\, n \in \mathbb{N}$$

Taking the limit in n completes the proof.

1.3.3 Failure of Memorylessness

We know from the proceeding section that any distribution on $\mathbb{R}_+$ other than the exponential distribution fails to be memoryless.

Here's an example that helps to clarify (although the support of the distribution is a proper subset of $\mathbb{R}_+$).

A random variable Y has the Pareto distribution with positive parameters t_0, α if

$$f(t) := \mathbb{P}\{Y > t\} = \begin{cases} 1 & \text{if } t \leq t_0 \\ (t_0/t)^\alpha & \text{if } t > t_0 \end{cases}$$

As a result, with $s > t_0$,

$$\mathbb{P}\{Y > s + t \mid Y > s\} = \frac{\mathbb{P}\{Y > s + t\}}{\mathbb{P}\{Y > s\}} = \left(\frac{t}{t + s}\right)^\alpha$$

Since this probability falls with s, the distribution is not memoryless.

If we have waited many hours for an event (i.e., s is large), then the probability of waiting another hour is relatively small.

1.4 Sums of Exponentials

A random variable W on $\mathbb{R}_+$ is said to have the Erlang distribution if its density has the form

$$f(t) = \frac{\lambda^n t^{n-1}}{(n-1)!} e^{-\lambda t} \qquad (t \geq 0)$$

for some $n \in \mathbb{N}$ and $\lambda > 0$.

The parameters n and λ are called the **shape** and **rate** parameters respectively.

The next figure shows the shape for two parameterizations.

```
<>:17: SyntaxWarning: invalid escape sequence '\l'
<>:17: SyntaxWarning: invalid escape sequence '\l'
/tmp/ipykernel_6500/1026889215.py:17: SyntaxWarning: invalid escape sequence '\l'
  ax.plot(t_grid, e(t_grid), label=f'$n={e.n}, \lambda={e.λ}$')
```
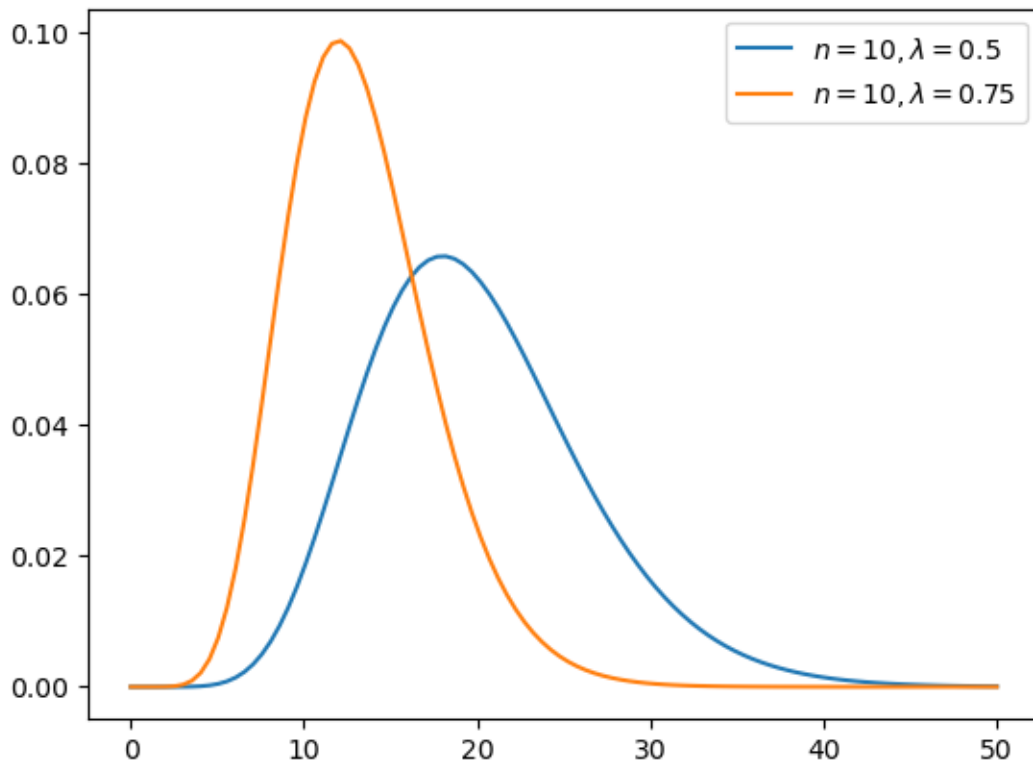

The CDF of the Erlang distribution is

$$F(t) = \mathbb{P}\{W \leq t\} = 1 - \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \tag{1.5}$$

The Erlang distribution is of interest to us because of the following fact.

Lemma 1.1 (Distribution of Sum of Exponentials)

If, for some $\lambda > 0$, the sequence (W_i) is IID and exponentially distributed with rate λ, then $J_n := \sum_{i=1}^n W_i$ has the Erlang distribution with shape n and rate λ.

This connects to Poisson process theory, as we shall soon see.

1.5 Exercises

Exercise 1.1

Due to its memoryless property, we can "stop" and "restart" an exponential draw without changing its distribution.

To illustrate this, we can think of fixing $\lambda > 0$, drawing from $\text{Exp}(\lambda)$, and stopping and restarting whenever a threshold s is crossed.

In particular, consider the random variable X defined as follows:

- Draw Y from $\text{Exp}(\lambda)$.
- If $Y \leq s$, set $X = Y$.
- If not, draw Z independently from $\text{Exp}(\lambda)$ and set $X = s + Z$.

Show that $X \sim \text{Exp}(\lambda)$.

Exercise 1.2

Fix $\lambda = 0.5$ and $s = 1.0$.

Simulate $1,000$ draws of X using the algorithm above.

Plot the fraction of the sample exceeding t for each $t \geq 0$ (on a grid) and compare to $t \mapsto e^{-\lambda t}$.

Is the fit good? How about if the number of draws is increased?

Are the results in line with those of the previous exercise?

1.6 Solutions

Note: code is currently not supported in `sphinx-exercise` so code-cell solutions are immediately after this solution block.

Solution to Exercise 1.1

Let X be constructed as in the statement of the exercise and fix $t > 0$.

Notice that $X > s + t$ if and only if $Y > s$ and $Z > t$.

As a result of this fact and independence,

$$\mathbb{P}\{X > s + t\} = \mathbb{P}\{Y > s\}\mathbb{P}\{Z > t\} = e^{-\lambda(s+t)}$$

At the same time, $X > s - t$ if and only if $Y > s - t$, so

$$\mathbb{P}\{X > s - t\} = \mathbb{P}\{Y > s - t\} = e^{-\lambda(s-t)}$$

Either way, we have $X \sim \text{Exp}(\lambda)$, as was to be shown.

Solution to Exercise 1.2

Here's one solution, starting with 1,000 draws.

```python
λ = 0.5
np.random.seed(1234)
t_grid = np.linspace(0, 10, 200)

@njit
def draw_X(s=1.0, n=1_000):
    draws = np.empty(n)
    for i in range(n):
        Y = np.random.exponential(scale=1/λ)
        if Y <= s:
            X = Y
        else:
            Z = np.random.exponential(scale=1/λ)
            X = s + Z
        draws[i] = X
    return draws

fig, ax = plt.subplots()
draws = draw_X()
empirical_exceedance = [np.mean(draws > t) for t in t_grid]
ax.plot(t_grid, np.exp(- λ * t_grid), label='exponential exceedance')
ax.plot(t_grid, empirical_exceedance, label='empirical exceedance')
ax.legend()

plt.show()
```
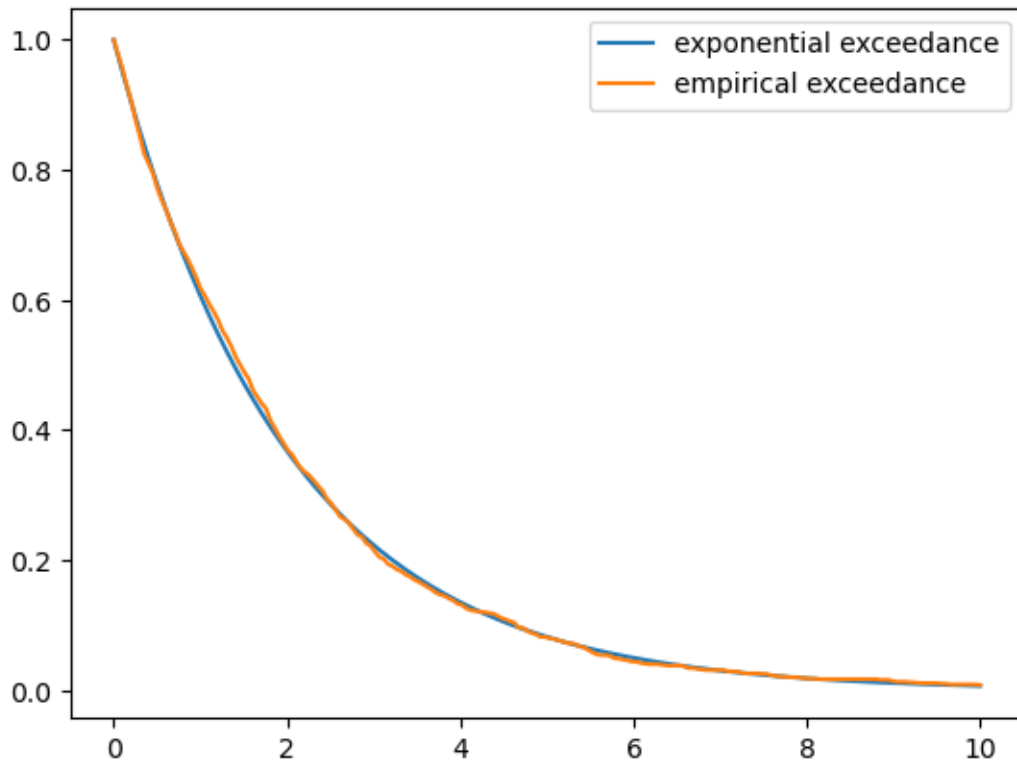
Solution to Exercise 1.2

Solution Continued:

The fit is already very close, which matches with the theory in Exercise 1.

The two lines become indistinguishable as n is increased further.

```python
fig, ax = plt.subplots()
draws = draw_X(n=10_000)
empirical_exceedance = [np.mean(draws > t) for t in t_grid]
ax.plot(t_grid, np.exp(- λ * t_grid), label='exponential exceedance')
ax.plot(t_grid, empirical_exceedance, label='empirical exceedance')
ax.legend()

plt.show()
```
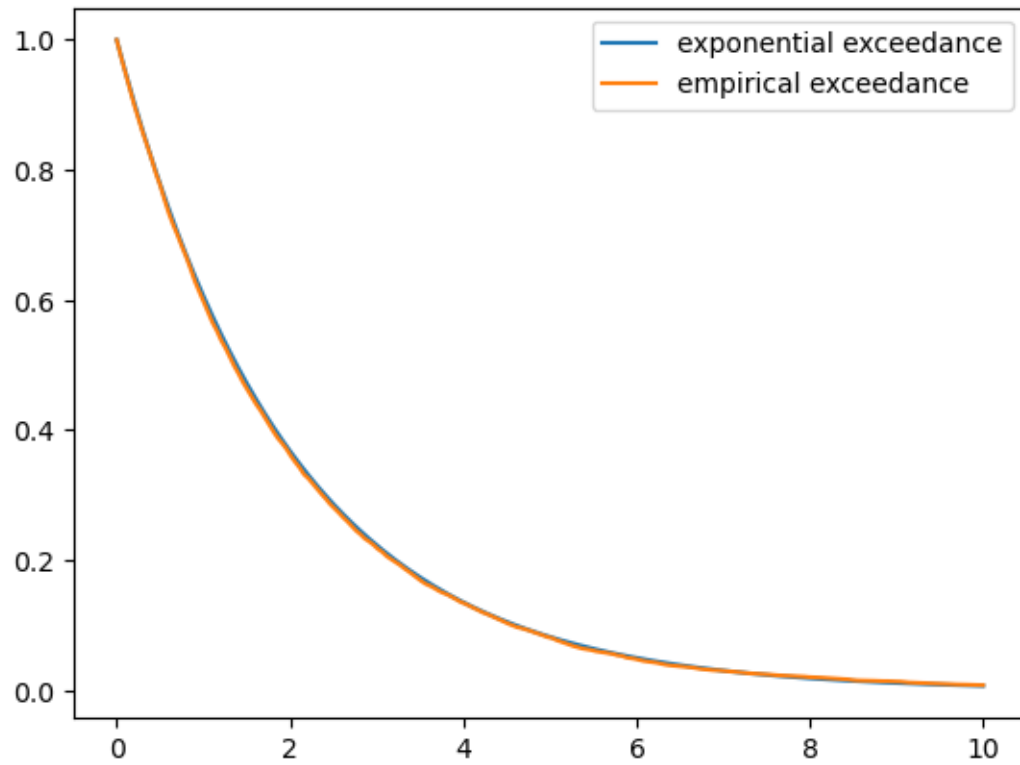
POISSON PROCESSES

In addition to what's in Anaconda, this lecture will need the following libraries:

```
!pip install quantecon
```

2.1 Overview

Counting processes count the number of "arrivals" occurring by a given time (e.g., the number of visitors to a website, the number of customers arriving at a restaurant, etc.)

Counting processes become Poisson processes when the time interval between arrivals is IID and exponentially distributed.

Exponential distributions and Poisson processes have deep connections to continuous time Markov chains.

For example, Poisson processes are one of the simplest nontrivial examples of a continuous time Markov chain.

In addition, when continuous time Markov chains jump between states, the time between jumps is *necessarily* exponentially distributed.

In discussing Poisson processes, we will use the following imports:

```python
import numpy as np
import matplotlib.pyplot as plt
import quantecon as qe
from numba import njit
from scipy.special import factorial, binom
```

2.2 Counting Processes

Let's start with the general case of an arbitrary counting process.

2.2.1 Jumps and Counts

Let (J_k) be an increasing sequence of nonnegative random variables satisfying $J_k \to \infty$ with probability one.

For example, J_k might be the time the k-th customer arrives at a shop.

Then

$$N_t := \sum_{k \geq 0} k\mathbb{1}\{J_k \leq t < J_{k+1}\} \tag{2.1}$$

is the number of customers that have visited by time t.

The next figure illustrate the definition of N_t for a given jump sequence $\{J_k\}$.
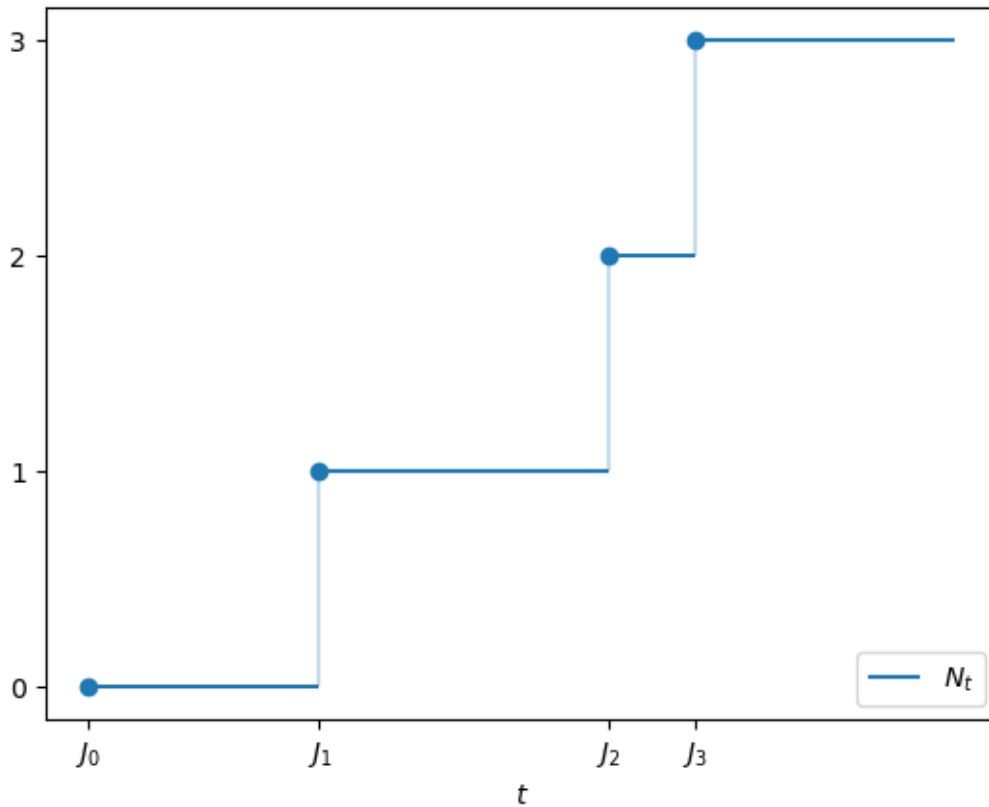
An alternative but equivalent definition is

$$N_t := \max\{k \geq 0 \mid J_k \leq t\}$$

As a function of t, the process N_t is called a **counting process**.

The jump times (J_k) are sometimes called **arrival times** and the intervals $J_k - J_{k-1}$ are called **wait times** or **holding times**.

2.2.2 Exponential Holding Times

A Poisson process is a counting process with independent exponential holding times.

In particular, suppose that the arrival times are given by $J_0 = 0$ and

$$J_k := W_1 + \cdots W_k$$

where (W_i) are IID exponential with some fixed rate λ.

Then the associated counting process (N_t) is called a **Poisson process** with rate λ.

The rationale behind the name is that, for each $t > 0$, the random variable N_t has the Poisson distribution with parameter $t\lambda$.

In other words,

$$\mathbb{P}\{N_t = k\} = e^{-t\lambda}\frac{(t\lambda)^k}{k!} \qquad (k = 0, 1, ...) \tag{2.2}$$

For example, since $N_t = 0$ if and only if $W_1 > t$, we have

$$\mathbb{P}\{N_t = 0\} = \mathbb{P}\{W_1 > t\} = e^{-t\lambda}$$

and the right hand side agrees with (2.2) when $k = 0$.

This sets up a proof by induction, which is time consuming but not difficult — the details can be found in §29 of [Howard, 2017].

Another way to show that N_t is Poisson with rate λ is to appeal to *Lemma 1.1*.

We observe that

$$\mathbb{P}\{N_t \leq n\} = \mathbb{P}\{J_{n+1} > t\} = 1 - \mathbb{P}\{J_{n+1} \leq t\}$$

Inserting the expression for the Erlang CDF in (1.5) with shape $n + 1$ and rate λ, we obtain

$$\mathbb{P}\{N_t \leq n\} = \sum_{k=0}^{n} \frac{(t\lambda)^k}{k!} e^{-t\lambda}$$

This is the (integer valued) CDF for the Poisson distribution with parameter $t\lambda$.

An exercise at the end of the lecture asks you to verify that N_t is Poisson-$(t\lambda)$ informally via simulation.

The next figure shows one realization of a Poisson process (N_t), with jumps at each new arrival.
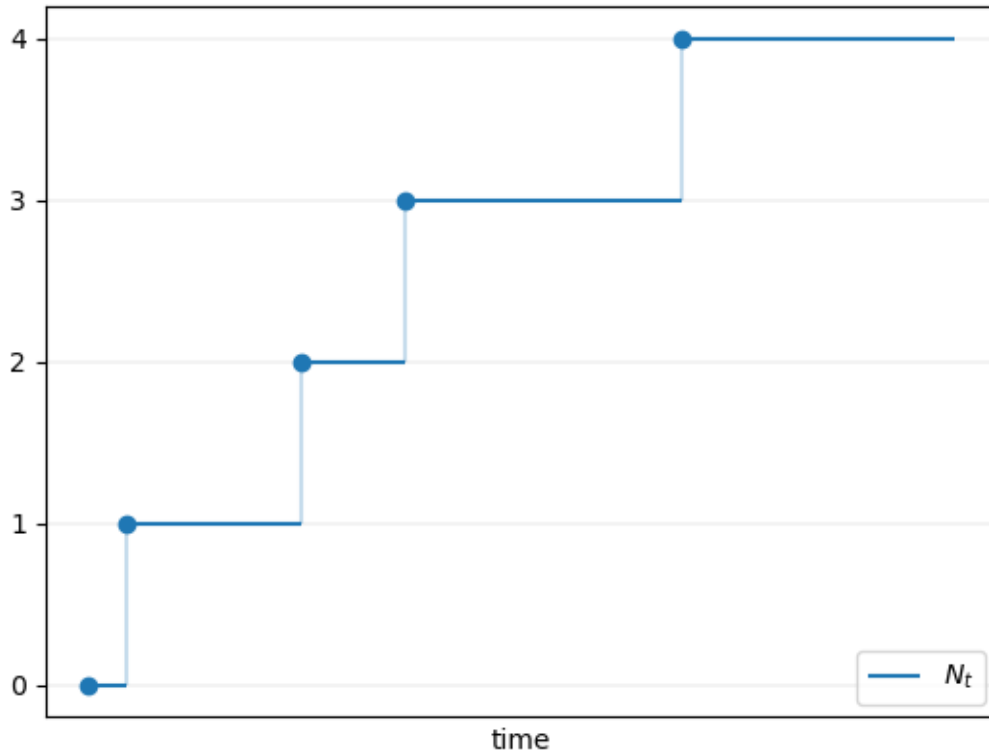
2.3 Stationary Independent Increments

One of the defining features of a Poisson process is that it has stationary and independent increments.

This is due to the memoryless property of exponentials.

It means that

1. the variables $\{N_{t_{i+1}} - N_{t_i}\}_{i \in I}$ are independent for any strictly increasing finite sequence $(t_i)_{i \in I}$ and

2. the distribution of $N_{t+h} - N_t$ depends on h but not t.

A detailed proof can be found in Theorem 2.4.3 of [Norris, 1998].

Instead of repeating this, we provide some intuition from a discrete approximation.

In the discussion below, we use the following well known fact: If (θ_n) is a sequence such that $n\theta_n$ converges, then

$$\text{Binomial}(n, \theta_n) \approx \text{Poisson}(n\theta_n) \quad \text{for large } n \tag{2.3}$$

(The exercises ask you to examine this claim visually.)

We now return to *the environment* where we linked the geometric distribution to the exponential.

That is, we fix small $h > 0$ and let $t_i := ih$ for all $i \in \mathbb{Z}_+$.

Let (V_i) be IID binary random variables with $\mathbb{P}\{V_i = 1\} = h\lambda$ for some $\lambda > 0$.

Linking to our previous discussion,

- either one or zero customers visits a shop at each t_i.

- $V_i = 1$ means that a customer visits at time t_i.

- Visits occur with probability $h\lambda$, which is proportional to the length of the interval between grid points.

We learned that the wait time until the first visit is approximately exponential with rate $t\lambda$.

Since (V_i) is IID, the same is true for the second wait time and so on.

Moreover, these wait times are independent, since they depend on separate subsets of (V_i).

Let $\hat{N}_t$ count the number of visits by time t, as shown in the next figure.

($V_i = 1$ is indicated by a vertical line at $t_i = ih$.)

```
<>:15: SyntaxWarning: invalid escape sequence '\h'
<>:15: SyntaxWarning: invalid escape sequence '\h'
/tmp/ipykernel_6548/1718099611.py:15: SyntaxWarning: invalid escape sequence '\h'
  ax.step(t_grid, np.insert(N, 0, 0)[:-1], label='$\hat N_t$')
```
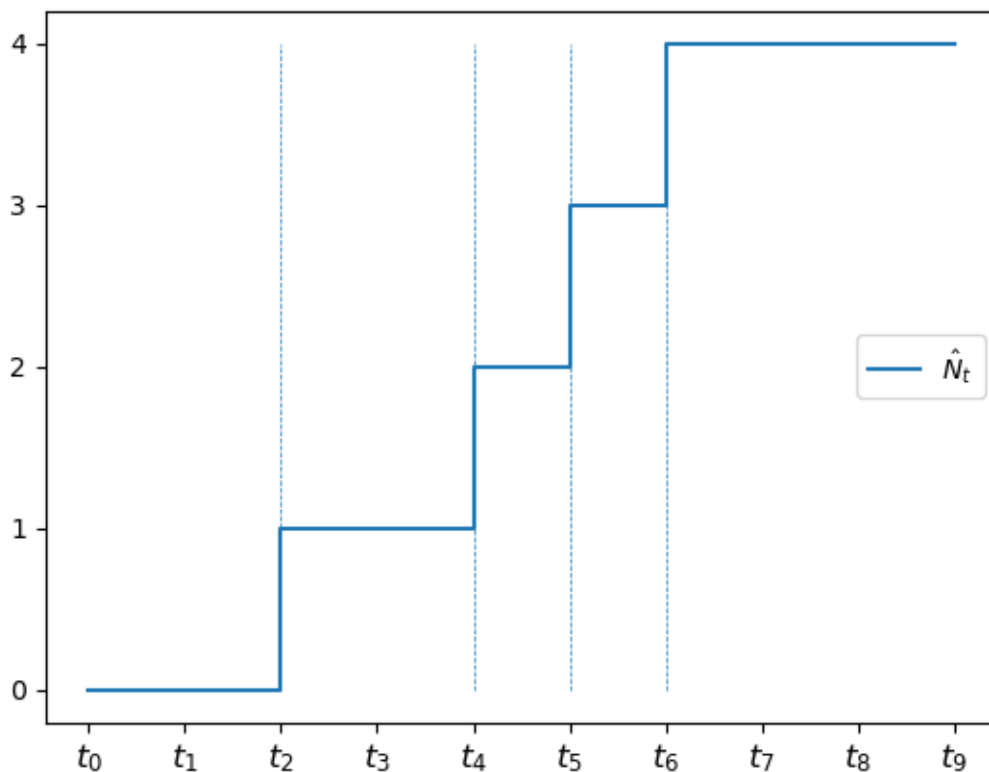

We expect from the discussion above that $(\hat{N}_t)$ approximates a Poisson process.

This intuition is correct because, fixing t, letting $k := \max\{i \in \mathbb{Z}_+ : t_i \leq t\}$ and applying (2.3), we have

$$\hat{N}_t = \sum_{i=1}^{k} V_i \sim \text{Binomial}(k, h\lambda) \approx \text{Poisson}(kh\lambda)$$

Using the fact that $kh = t_k \approx t$ as $h \to 0$, we see that $\hat{N}_t$ is approximately Poisson with rate $t\lambda$, just as we expected.

This approximate construction of a Poisson process helps illustrate the property of stationary independent increments.

For example, if we fix s, t, then $\hat{N}_{s+t} - \hat{N}_s$ is the number of visits between s and $s + t$, so that

$$\hat{N}_{s+t} - \hat{N}_s = \sum_{i} V_i \mathbb{1}\{s \leq t_i < s + t\}$$

Suppose there are k grid points between s and $s + t$, so that $t \approx kh$.

Then

$$\hat{N}_{s+t} - \hat{N}_s \sim \text{Binomial}(k, h\lambda) \approx \text{Poisson}(kh\lambda) \approx \text{Poisson}(t\lambda)$$

This illustrates the idea that, for a Poisson process (N_t), we have

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

In particular, increments are stationary (the distribution depends on t but not s).

The approximation also illustrates independence of increments, since, in the approximation, increments depend on separate subsets of (V_i).

2.4 Uniqueness

What other counting processes have stationary independent increments?

Remarkably, the answer is none:

Theorem 2.1 (Characterization of Poisson Processes)

If (M_t) is a stochastic process supported on $\mathbb{Z}_+$ and starting at 0 with the property that its increments are stationary and independent, then (M_t) is a Poisson process.

In particular, there exists a $\lambda > 0$ such that

$$M_{s+t} - M_s \sim \text{Poisson}(t\lambda)$$

for any s, t.

The proof is similar to our earlier proof that the exponential distribution is the only memoryless distribution.

Details can be found in Section 6.2 of [Pardoux, 2008] or Theorem 2.4.3 of [Norris, 1998].

2.4.1 The Restarting Property

An important consequence of stationary independent increments is the restarting property, which means that, when simulating, we can freely stop and restart a Poisson process at any time:

Theorem 2.2 (Poisson Processes can be Paused and Restarted)

If (N_t) is a Poisson process, $s > 0$ and (M_t) is defined by $M_t = N_{s+t} - N_s$ for $t \geq 0$, then (M_t) is a Poisson process independent of $(N_r)_{r \leq s}$.

 Proof. Independence of (M_t) and $(N_r)_{r \leq s}$ follows from indepence of the increments of (N_t).

In view of the uniqueness statement above, we can verify that (M_t) is a Poisson process by showing that (M_t) starts at zero, takes values in $\mathbb{Z}_+$ and has stationary independent increments.

It is clear that (M_t) starts at zero and takes values in $\mathbb{Z}_+$.

In addition, if we take any $t < t'$, then

$$M_{t'} - M_t = N_{s+t'} - N_{s+t} \sim \text{Poisson}((t'-t)\lambda)$$

Hence (M_t) has stationary increments and, using the relation $M_{t'} - M_t = N_{s+t'} - N_{s+t}$ again, the increments are independent as well.

We conclude that $(N_{s+t} - N_s)_{t \geq 0}$ is indeed a Poisson process independent of $(N_r)_{r \leq s}$.

2.5 Exercises

Exercise 2.1

Fix $\lambda > 0$ and draw $\{W_i\}$ as IID exponentials with rate λ.

Set $J_n := W_1 + \cdots W_n$ with $J_0 = 0$ and $N_t := \sum_{n \geq 0} n \mathbb{1}\{J_n \leq t < J_{n+1}\}$.

Provide a visual test of the claim that N_t is Poisson with parameter $t\lambda$.

Do this by fixing $t = T$, generating many independent draws of N_T and comparing the empirical distribution of the sample with a Poisson distribution with rate $T\lambda$.

Try first with $\lambda = 0.5$ and $T = 10$.

Exercise 2.2

In the lecture we used the fact that $\text{Binomial}(n, \theta) \approx \text{Poisson}(n\theta)$ when n is large and θ is small.

Investigate this relationship by plotting the distributions side by side.

Experiment with different values of n and θ.

2.6 Solutions

Note: code is currently not supported in `sphinx-exercise` so code-cell solutions are immediately after this solution block.

Solution to Exercise 2.1

Here is one solution.

The figure shows that the fit is already good with a modest sample size.

Increasing the sample size will further improve the fit.

```python
λ = 0.5
T = 10

def poisson(k, r):
    "Poisson pmf with rate r."
    return np.exp(-r) * (r**k) / factorial(k)

@njit
def draw_Nt(max_iter=1e5):
    J = 0
    n = 0
    while n < max_iter:
        W = np.random.exponential(scale=1/λ)
        J += W
        if J > T:
            return n
        n += 1

@njit
def draw_Nt_sample(num_draws):
    draws = np.empty(num_draws)
    for i in range(num_draws):
        draws[i] = draw_Nt()
    return draws


sample_size = 10_000
sample = draw_Nt_sample(sample_size)
max_val = sample.max()
vals = np.arange(0, max_val+1)

fig, ax = plt.subplots()

ax.plot(vals, [poisson(v, T * λ) for v in vals],
    marker='o', label='poisson')
ax.plot(vals, [np.mean(sample==v) for v in vals],
    marker='o', label='empirical')

ax.legend(fontsize=12)
plt.show()
```
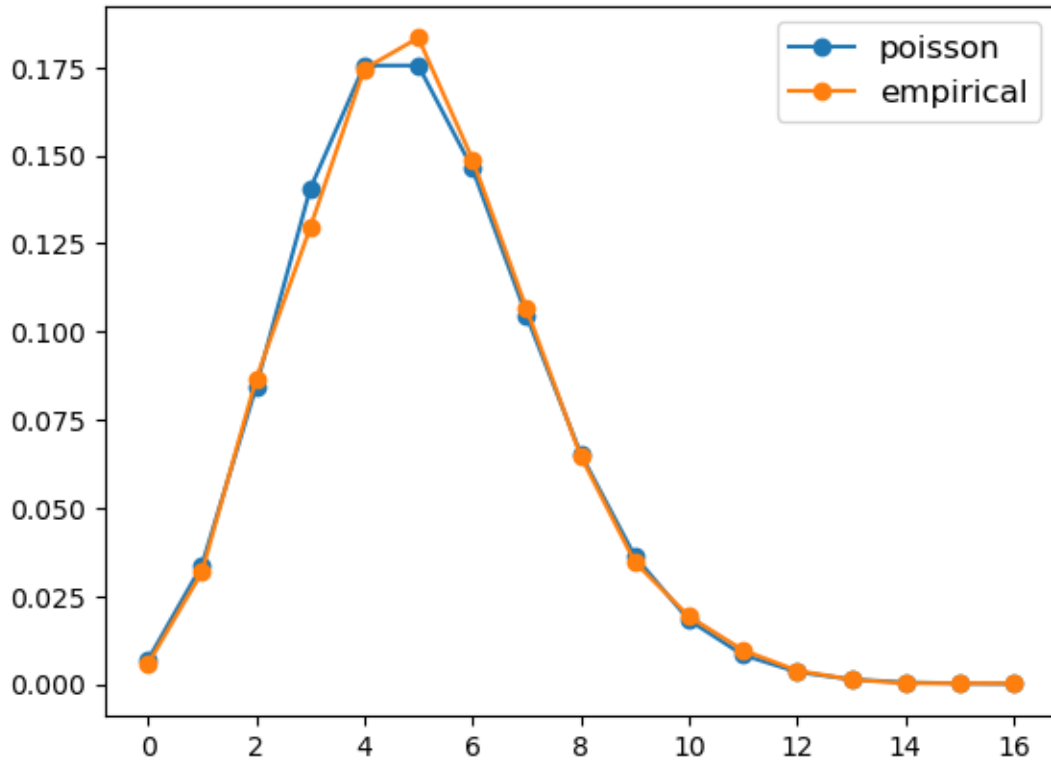
Solution to Exercise 2.2

Here is one solution. It shows that the approximation is good when n is large and θ is small.

```python
def binomial(k, n, p):
    # Binomial(n, p) pmf evaluated at k
    return binom(n, k) * p**k * (1-p)**(n-k)


θ_vals = 0.5, 0.2, 0.1

n_vals = 50, 75, 100

fig, axes = plt.subplots(len(n_vals), 1, figsize=(6, 12))

for n, θ, ax in zip(n_vals, θ_vals, axes.flatten()):

    k_grid = np.arange(n)
    binom_vals = [binomial(k, n, θ) for k in k_grid]
    poisson_vals = [poisson(k, n * θ) for k in k_grid]
    ax.plot(k_grid, binom_vals, 'o-', alpha=0.5, label='binomial')
    ax.plot(k_grid, poisson_vals, 'o-', alpha=0.5, label='Poisson')
    ax.set_title(f'$n={n}$ and $\\theta = {θ}$')
    ax.legend(fontsize=12)

fig.tight_layout()
plt.show()
```
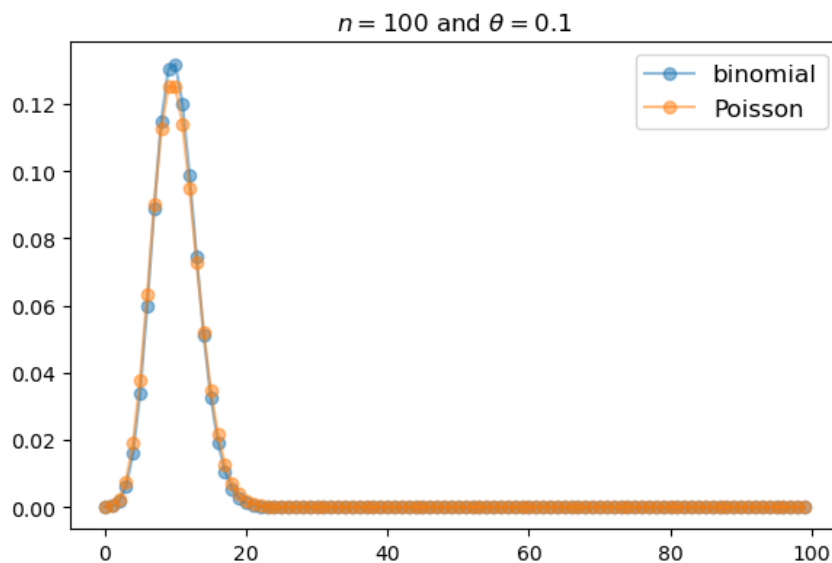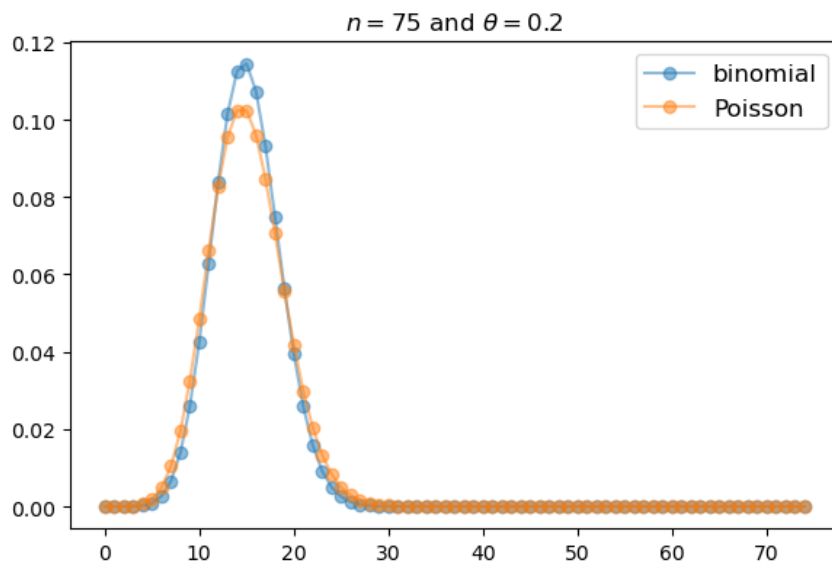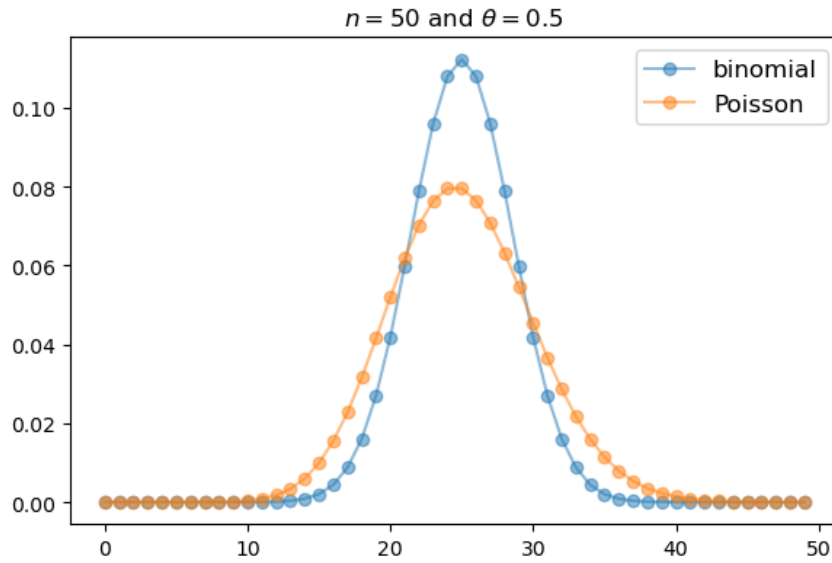
THE MARKOV PROPERTY

In addition to what's in Anaconda, this lecture will need the following libraries:

```
!pip install quantecon
```

3.1 Overview

A continuous time stochastic process is said to have the Markov property if its past and future are independent given the current state.

(A more formal definition is provided below.)

As we will see, the Markov property imposes a large amount of structure on continuous time processes.

This structure leads to elegant and powerful results on evolution and dynamics.

At the same time, the Markov property is general enough to cover many applied problems, as described in *the introduction*.

3.1.1 Setting

In this lecture, the state space where dynamics evolve will be a countable set, denoted henceforth by S, with typical elements x, y.

(Note that "countable" is understood to include finite.)

Regarding notation, in what follows, $\sum_{x \in S}$ is abbreviated to $\sum_x$, the supremum $\sup_{x \in S}$ is abbreviated to $\sup_x$ and so on.

A **distribution** on S is a function ϕ from S to $\mathbb{R}_+$ with $\sum_x \phi(x) = 1$.

Let $\mathcal{D}$ denote the set of all distributions on S.

To economize on terminology, we define a **matrix** A on S to be a map from $S \times S$ to $\mathbb{R}$.

When S is finite, this reduces to the usual notion of a matrix, and, whenever you see expressions such as $A(x, y)$ below, you can mentally identify them with more familiar matrix notation, such as A_{ij}, if you wish.

The product of two matrices A and B is defined by

$$(AB)(x, y) = \sum_z A(x, z)B(z, y) \qquad ((x, y) \in S \times S) \tag{3.1}$$

If S is finite, then this is just ordinary matrix multiplication.

In statements involving matrix algebra, we *always treat distributions as row vectors*, so that, for $\phi \in \mathcal{D}$ and given matrix A,

$$(\phi A)(y) = \sum_x \phi(x) A(x, y)$$

We will use the following imports

```python
import numpy as np
import scipy as sp
import matplotlib.pyplot as plt

import quantecon as qe
from numba import njit

from scipy.linalg import expm
from scipy.stats import binom

from matplotlib import cm
from mpl_toolkits.mplot3d import Axes3D
```

3.2 Markov Processes

We now introduce the definition of Markov processes, first reviewing the discrete case and then shifting to continuous time.

3.2.1 Discrete Time, Finite State

The simplest Markov processes are those with a discrete time parameter and finite state space.

Assume for now that S has n elements and let P be a **Markov matrix**, which means that $P(x, y) \geq 0$ and $\sum_y P(x, y) = 1$ for all x.

In applications, $P(x, y)$ represents the probability of transitioning from x to y in one step.

A **Markov chain** $(X_t)_{t \in \mathbb{Z}_+}$ on S with Markov matrix P is a sequence of random variables satisfying

$$\mathbb{P}\{X_{t+1} = y \mid X_0, X_1, \dots, X_t\} = P(X_t, y) \tag{3.2}$$

with probability one for all $y \in S$ and any $t \in \mathbb{Z}_+$.

In addition to connecting probabilities to the Markov matrix, (3.2) says that the process depends on its history only through the current state.

We recall that, if X_t has distribution ϕ, then X_{t+1} has distribution ϕP.

Since ϕ is understood as a row vector, the meaning is

$$(\phi P)(y) = \sum_x \phi(x) P(x, y) \qquad (y \in S) \tag{3.3}$$

The Joint Distribution

In general, for given Markov matrix P, there can be many Markov chains (X_t) that satisfy (3.2).

This is due to the more general observation that, for a given distribution ϕ, we can construct many random variables having distribution ϕ.

(The exercises below ask for one example.)

Hence P is, in a sense, a more primitive object than (X_t).

There is another way to see the fundamental importance of P, which is by constructing the joint distribution of (X_t) from P.

Let S^∞ represent the space of S-valued sequences $(x_0, x_1, x_2, ...)$.

Fix an initial condition $\psi \in \mathcal{D}$ and a Markov matrix P on S.

The **joint distribution** of a Markov chain (X_t) satisfying (3.2) and $X_0 \sim \psi$ is the distribution $\mathbf{P}_\psi$ over S^∞ such that

$$\mathbb{P}\{X_{t_1} = y_1, ..., X_{t_m} = y_m\} = \mathbf{P}_\psi\{(x_t) \in S^\infty : x_{t_i} = y_i \text{ for } i = 1, ... m\} \tag{3.4}$$

for any m positive integers t_i and m elements y_i of the state space S.

(Joint distributions of discrete time processes are uniquely defined by their values at finite collections of times — see, for example, Theorem 7.2 of [Walsh, 2012].)

We can construct $\mathbf{P}_\psi$ by first defining P_ψ^n over the finite Cartesian product S^{n+1} via

$$\mathbf{P}_\psi^n(x_0, x_1, ..., x_n) = \psi(x_0)P(x_0, x_1) \times \cdots \times P(x_{n-1}, x_n) \tag{3.5}$$

For any Markov chain (X_t) satisfying (3.2) and $X_0 \sim \psi$, the restriction $(X_0, ..., X_n)$ has joint distribution $\mathbf{P}_\psi^n$.

This is a solved exercise below.

The last step is to show that the family $(\mathbf{P}_\psi^n)$ defined at each $n \in \mathbb{N}$ extends uniquely to a distribution $\mathbf{P}_\psi$ over the infinite sequences in S^∞.

That this is true follows from a well known theorem of Kolmogorov.

Hence P defines the joint distribution $\mathbf{P}_\psi$ when paired with any initial condition ψ.

3.2.2 Extending to Infinite (Countable) State Spaces

When S is infinite, the same idea carries through.

Consistent with the finite case, a **Markov matrix** is a map P from $S \times S$ to $\mathbb{R}_+$ satisfying

$$\sum_y P(x, y) = 1 \text{ for all } x \in S$$

The definition of a Markov chain $(X_t)_{t \in \mathbb{Z}_+}$ on S with Markov matrix P is exactly as in (3.2).

Given Markov matrix P and $\phi \in \mathcal{D}$, we define ϕP by (3.3).

Then, as before, ϕP can be understood as the distribution of X_{t+1} when X_t has distribution ϕ.

The function ϕP is in $\mathcal{D}$, since, by (3.3), it is nonnegative and

$$\sum_y (\phi P)(y) = \sum_y \sum_x P(x, y)\phi(x) = \sum_x \sum_y P(x, y)\phi(x) = \sum_x \phi(x) = 1$$

(Swapping the order of infinite sums is justified here by the fact that all elements are nonnegative — a version of Tonelli's theorem).

If P and Q are Markov matrices on S, then, using the definition in (3.1),

$$(PQ)(x, y) := \sum_z P(x, z)Q(z, y)$$

It is not difficult to check that PQ is again a Markov matrix on S.

The elements of P^k, the k-th product of P with itself, give k step transition probabilities.

For example, we have

$$P^k(x, y) = (P^{k-j}P^j)(x, y) = \sum_z P^{k-j}(x, z)P^j(z, y) \tag{3.6}$$

which is a version of the (discrete time) Chapman-Kolmogorov equation.

Equation (3.6) can be obtained from the law of total probability: if (X_t) is a Markov chain with Markov matrix P and initial condition $X_0 = x$, then

$$\mathbb{P}\{X_k = y\} = \sum_z \mathbb{P}\{X_k = y \mid X_j = z\}\mathbb{P}\{X_j = z\}$$

All of the *preceding discussion* on the connection between P and the joint distribution of (X_t) when S is finite carries over to the current setting.

3.2.3 The Continuous Time Case

A **continuous time stochastic process** on S is a collection (X_t) of S-valued random variables X_t defined on a common probability space and indexed by $t \in \mathbb{R}_+$.

Let I be the Markov matrix on S defined by $I(x, y) = \mathbb{1}\{x = y\}$.

A **Markov semigroup** is a family (P_t) of Markov matrices on S satisfying

1. $P_0 = I$,

2. $\lim_{t \to 0} P_t(x, y) = I(x, y)$ for all x, y in S, and

3. the semigroup property $P_{s+t} = P_s P_t$ for all $s, t \geq 0$.

The interpretation of $P_t(x, y)$ is the probability of moving from state x to state y in t units of time.

As such it is natural that $P_0(x, y) = 1$ if $x = y$ and zero otherwise, which is condition 1.

Condition 2 is continuity with respect to t, which might seem restrictive but it is in fact very mild.

For all practical applications, probabilities do not jump — although the chain (X_t) itself can of course jump from state to state as time goes by.[1]

The semigroup property in condition 3 is nothing more than a continuous time version of the Chapman-Kolmogorov equation.

This becomes clearer if we write it more explicitly as

$$P_{s+t}(x, y) = \sum_z P_s(x, z)P_t(z, y) \tag{3.7}$$

A stochastic process (X_t) is called a (time homogeneous) **continuous time Markov chain** on S with Markov semigroup (P_t) if

$$\mathbb{P}\{X_{s+t} = y \mid \mathcal{F}_s\} = P_t(X_s, y) \tag{3.8}$$

---

[1] On a technical level, right continuity of paths for (X_t) implies condition 2, as proved in Theorem 2.12 of [Liggett, 2010]. Right continuity of paths allows for jumps, but insists on only finitely many jumps in any bounded interval.

with probability one for all $y \in S$ and $s, t \geq 0$.

Here $\mathscr{F}_s$ is the history $(X_r)_{r \leq s}$ of the process up until time s.

If you are an economist you might call $\mathscr{F}_s$ the "information set" at time s.

If you are familiar with measure theory, you can understand $\mathscr{F}_s$ as the σ-algebra generated by $(X_r)_{r \leq s}$.

Analogous to the discrete time case, the joint distribution of (X_t) is determined by its Markov semigroup plus an initial condition.

This distribution is defined over the set of all right continuous functions $\mathbb{R}_+ \ni t \mapsto x_t \in S$, which we call rcS.

Next one builds finite dimensional distributions over rcS using expressions similar to (3.5).

Finally, the Kolmogorov extension theorem is applied, similar to the discrete time case.

Corollary 6.4 of [Le Gall, 2016] provides full details.

3.2.4 The Canonical Chain

Given a Markov semigroup (P_t) on S, does there always exist a continuous time Markov chain (X_t) such that (3.8) holds?

The answer is affirmative.

To illustrate, pick any Markov semigroup (P_t) on S and fix initial condition ψ.

Next, create the corresponding joint distribution $\mathbf{P}_\psi$ over rcS, as described above.

Now, for each $t \geq 0$, let π_t be the time t projection on rcS, which maps any right continuous function (x_τ) into its time t value x_t.

Finally, let X_t be an S-valued function on rcS defined at $(x_\tau) \in rcS$ by $\pi_t((x_\tau))$.

In other words, after $\mathbf{P}_\psi$ picks out some time path $(x_\tau) \in rcS$, the Markov chain (X_t) simply reports this time path.

Hence (X_t) automatically has the correct distribution.

The chain (X_t) constructed in this way is called the **canonical chain** for the semigroup (P_t) and initial condition ψ.

3.2.5 Simulation and Probabilistic Constructions

While we have answered the existence question in the affirmative, the canonical construction is quite abstract.

Moreover, there is little information about how we might simulate such a chain.

Fortunately, it turns out that there are more concrete ways to build continuous time Markov chains from the objects that describe their distributions.

We will learn about these in a *later lecture*.

3.3 Implications of the Markov Property

The Markov property carries some strong implications that are not immediately obvious.

Let's take some time to explore them.

3.3.1 Example: Failure of the Markov Property

Let's look at how the Markov property can fail, via an intuitive rather than formal discussion.

Let (X_t) be a continuous time stochastic process with state space $S = \{0, 1\}$.

The process starts at 0 and updates at follows:

1. Draw W independently from a fixed Pareto distribution.

2. Hold (X_t) in its current state for W units of time and then switch to the other state.

3. Go to step 1.

What is the probability that $X_{s+h} = i$ given both the history $(X_r)_{r \leq s}$ and current information $X_s = i$?

If h is small, then this is close to the probability that there are zero switches over the time interval $(s, s+h]$.

To calculate this probability, it would be helpful to know how long the state has been at current state i.

This is because the Pareto distribution *is not memoryless*.

(With a Pareto distribution, if we know that X_t has been at i for a long time, then a switch in the near future becomes more likely.)

As a result, the history prior to X_s is useful for predicting X_{s+h}, even when we know X_s.

Thus, the Markov property fails.

3.3.2 Restrictions Imposed by the Markov Property

From the discussion above, we see that, for continuous time Markov chains, the length of time between jumps must be memoryless.

Recall that, by *Theorem 1.1*, the only memoryless distribution supported on $\mathbb{R}_+$ is the exponential distribution.

Hence, a continuous time Markov chain waits at states for an exponential amount of time and then jumps.

The way that the new state is chosen must also satisfy the Markov property, which adds another restriction.

In summary, we already understand the following about continuous time Markov chains:

1. Holding times are independent exponential draws.

2. New states are chosen in a ``Markovian'' way, independent of the past given the current state.

We just need to clarify the details in these steps to have a complete description.

3.4 Examples of Markov Processes

Let's look at some examples of processes that possess the Markov property.

3.4.1 Example: Poisson Processes

The Poisson process discussed in our *previous lecture* is a Markov process on state space $\mathbb{Z}_+$.

To obtain the Markov semigroup, we observe that, for $k \geq j$,

$$\mathbb{P}\{N_{s+t} = k \mid N_s = j\} = \mathbb{P}\{N_{s+t} - N_s = k - j \mid N_s = j\} = \mathbb{P}\{N_{s+t} - N_s = k - j\}$$

where the last step is due to independence of increments.

From stationarity of increments we have

$$\mathbb{P}\{N_{s+t} - N_s = k - j\} = \mathbb{P}\{N_t = k - j\} = e^{-\lambda t}\frac{(\lambda t)^{k-j}}{(k-j)!}$$

In summary, the Markov semigroup is

$$P_t(j, k) = e^{-\lambda t}\frac{(\lambda t)^{k-j}}{(k-j)!} \tag{3.9}$$

whenever $j \leq k$ and $P_t(j, k) = 0$ otherwise.

This chain of equalities was obtained with $N_s = j$ for arbitrary j, so we can replace j with N_s in (3.9) to verify the Markov property (3.8) for the Poisson process.

Under (3.9), each P_t is a Markov matrix and (P_t) is a Markov semigroup.

The proof of the semigroup property is a solved exercise below.[2]

3.5 A Model of Inventory Dynamics

Let X_t be the inventory of a firm at time t, taking values in the integers $0, 1, \ldots, b$.

If $X_t > 0$, then a customer arrives after W units of time, where $W \sim \mathrm{Exp}(\lambda)$ for some fixed $\lambda > 0$.

Upon arrival, each customer purchases $\min\{U, X_t\}$ units, where U is an IID draw from the geometric distribution started at 1 rather than 0:

$$\mathbb{P}\{U = k\} = (1 - \alpha)^{k-1}\alpha \qquad (k = 1, 2, \ldots, \ \alpha \in (0, 1))$$

If $X_t = 0$, then no customers arrive and the firm places an order for b units.

The order arrives after a delay of D units of time, where $D \sim \mathrm{Exp}(\lambda)$.

(We use the same λ here just for convenience, to simplify the exposition.)

---

[2] In the definition of P_t in (3.9), we use the convention that $0^0 = 1$, which leads to $P_0 = I$ and $\lim_{t \to 0} P_t(j, k) = I(j, k)$ for all j, k. These facts, along with the semigroup property, imply that (P_t) is a valid Markov semigroup.

3.5.1 Representation

The inventory process jumps to a new value either when a new customer arrives or when new stock arrives.

Between these arrival times it is constant.

Hence, to track X_t, it is enough to track the jump times and the new values taken at the jumps.

In what follows, we denote the jump times by $\{J_k\}$ and the values at jumps by $\{Y_k\}$.

Then we construct the state process via

$$X_t = \sum_{k \geq 0} Y_k \mathbb{1}\{J_k \leq t < J_{k+1}\} \qquad (t \geq 0) \tag{3.10}$$

3.5.2 Simulation

Let's simulate this process, starting at $X_0 = 0$.

As above,

- J_k is the time of the k-th jump (up or down) in inventory.
- Y_k is the size of the inventory after the k-th jump.
- (X_t) is defined from these objects via (3.10).

Here's a function that generates and returns one path $t \mapsto X_t$.

(We are not aiming for computational efficiency at this stage.)

```python
def sim_path(T=10, seed=123, λ=0.5, α=0.7, b=10):
    """
    Generate a path for inventory starting at b, up to time T.

    Return the path as a function X(t) constructed from (J_k) and (Y_k).
    """

    J, Y = 0, b
    J_vals, Y_vals = [J], [Y]
    np.random.seed(seed)

    while True:
        W = np.random.exponential(scale=1/λ)   # W ~ Exp(λ)
        J += W
        J_vals.append(J)
        if J >= T:
            break
        # Update Y
        if Y == 0:
            Y = b
        else:
            U = np.random.geometric(α)
            Y = Y - min(Y, U)
        Y_vals.append(Y)

    Y_vals = np.array(Y_vals)
    J_vals = np.array(J_vals)

    def X(t):
```

```
        if t == 0.0:
            return Y_vals[0]
        else:
            k = np.searchsorted(J_vals, t)
            return Y_vals[k-1]

    return X
```

Let's plot the process (X_t).
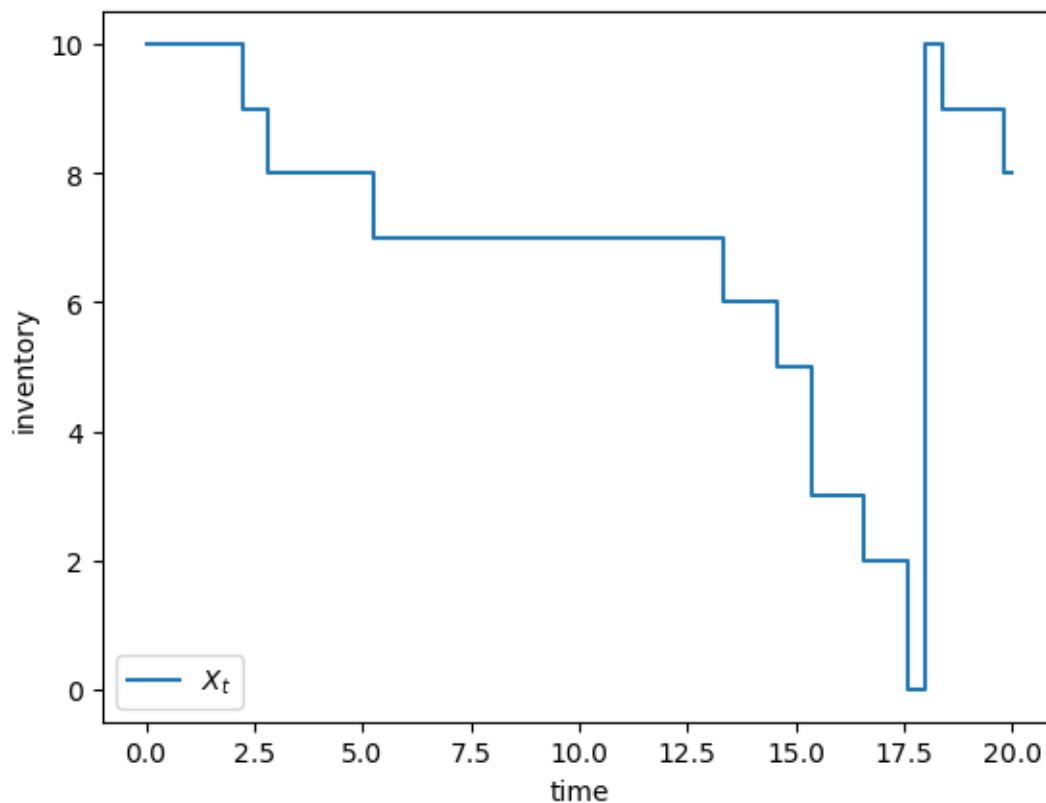
```
T = 20
X = sim_path(T=T)

grid = np.linspace(0, T, 100)

fig, ax = plt.subplots()
ax.step(grid, [X(t) for t in grid], label="$X_t$")

ax.set(xlabel="time", ylabel="inventory")

ax.legend()
plt.show()
```


As expected, inventory falls and then jumps back up to b.

3.5.3 The Embedded Jump Chain

In models such as the one described above, the embedded discrete time process (Y_k) is called the "embedded jump chain".

It is easy to see that (Y_k) is discrete time finite state Markov chain.

Its Markov matrix K is given by $K(x, y) = \mathbb{1}\{y = b\}$ when $x = 0$ and, when $0 < x \leq b$,

$$K(x, y) = \begin{cases} 0 & \text{if } y \geq x \\ \mathbb{P}\{x - U = y\} = (1 - \alpha)^{x-y-1}\alpha & \text{if } 0 < y < x \\ \mathbb{P}\{U \geq x\} = (1 - \alpha)^{x-1} & \text{if } y = 0 \end{cases} \tag{3.11}$$

3.5.4 Markov Property

The inventory model just described has the Markov property precisely because

1. the jump chain (Y_k) is Markov in discrete time and

2. the holding times are independent exponential draws.

Rather than providing more details on these points here, let us first describe a more general setting where the arguments will be clearer and more useful.

3.6 Jump Processes with Constant Rates

The examples we have focused on so far are special cases of Markov processes with constant jump intensities.

These processes turn out to be very representative (although the constant jump intensity will later be relaxed).

Let's now summarize the model and its properties.

3.6.1 Construction

The data for a Markov process on S with constant jump rates are

- a parameter $\lambda > 0$ called the **jump rate**, which governs the jump intensities and

- a Markov matrix K on S, called the **jump matrix**.

To run the process we also need an initial condition $\psi \in \mathcal{D}$.

The process (X_t) is constructed by holding at each state for an exponential amount of time, with rate λ, and then updating to a new state via K.

In more detail, the construction is

Algorithm 3.1 (Constant Rate Jump Chain)

Inputs $\psi \in \mathcal{D}$, positive constant λ, Markov matrix K

Outputs Markov chain (X_t)

1. draw Y_0 from ψ

2. set $k = 1$ and $J_0 = 0$

3. draw W_k from $\text{Exp}(\lambda)$ and set $J_k = J_{k-1} + W_k$

---

4. set $X_t = Y_{k-1}$ for all t such that $J_{k-1} \leq t < J_k$.

5. draw Y_k from $K(Y_{k-1}, \cdot)$

6. set $k = k + 1$ and go to step 3.

An alternative, more parsimonious way to express the same process is to take

- (N_t) to be a Poisson process with rate λ and

- (Y_k) to be a discrete time Markov chain with Markov matrix K

and then set

$$X_t := Y_{N_t} \text{ for all } t \geq 0$$

As before, the discrete time process (Y_k) is called the **embedded jump chain**.

(Not to be confused with (X_t), which is often called a "jump process" or "jump chain" due to the fact that it changes states with jumps.)

The draws (W_k) are called the **wait times** or **holding times**.

3.6.2 Examples

The Poisson process with rate λ is a jump process on $S = \mathbb{Z}_+$.

The holding times are obviously exponential with constant rate λ.

The jump matrix is just $K(i,j) = \mathbb{1}\{j = i + 1\}$, so that the state jumps up by one at every J_k.

The inventory model is also a jump process with constant rate λ, this time on $S = \{0, 1, \ldots, b\}$.

The jump matrix was given in (3.11).

3.6.3 Markov Property

Let's show that the jump process (X_t) constructed above satisfies the Markov property, and obtain the Markov semigroup at the same time.

We will use two facts:

- the jump chain (Y_k) has the Markov property in discrete time and

- the Poisson process has stationary independent increments.

From these facts it is intuitive that the distribution of X_{t+s} given the whole history $\mathcal{F}_s = \{(N_r)_{r \leq s}, (Y_k)_{k \leq N_s}\}$ depends only on X_s.

Indeed, if we know X_s, then we can simply

- *restart* the Poisson process from N_s and then

- starting from $X_s = Y_{N_s}$, update the embedded jump chain (Y_k) using K each time a new jump occurs.

Let's write this more mathematically.

Fixing $y \in S$ and $s, t \geq 0$, we have

$$\mathbb{P}\{X_{s+t} = y \mid \mathcal{F}_s\} = \mathbb{P}\{Y_{N_{s+t}} = y \mid \mathcal{F}_s\} = \mathbb{P}\{Y_{N_s + N_{s+t} - N_s} = y \mid \mathcal{F}_s\}$$

Recalling that $N_{s+t} - N_s$ is Poisson distributed with rate $t\lambda$, independent of the history $\mathcal{F}_s$, we can write the display above as

$$\mathbb{P}\{X_{s+t} = y \mid \mathcal{F}_s\} = \sum_{k \geq 0} \mathbb{P}\{Y_{N_s+k} = y \mid \mathcal{F}_s\} \frac{(t\lambda)^k}{k!} e^{-t\lambda}$$

Because the embedded jump chain is Markov with Markov matrix K, we can simplify further to

$$\mathbb{P}\{X_{s+t} = y \mid \mathcal{F}_s\} = \sum_{k \geq 0} K^k(Y_{N_s}, y) \frac{(t\lambda)^k}{k!} e^{-t\lambda} = \sum_{k \geq 0} K^k(X_s, y) \frac{(t\lambda)^k}{k!} e^{-t\lambda}$$

Since the expression above depends only on X_s, we have proved that (X_t) has the Markov property.

3.6.4 Transition Semigroup

The Markov semigroup can be obtained from our final result, conditioning on $X_s = x$ to get

$$P_t(x, y) = \mathbb{P}\{X_{s+t} = y \mid X_s = x\} = e^{-t\lambda} \sum_{k \geq 0} K^k(x, y) \frac{(t\lambda)^k}{k!}$$

If S is finite, we can write this in matrix form and use the definition of the matrix exponential to get

$$P_t = e^{-t\lambda} \sum_{k \geq 0} \frac{(t\lambda K)^k}{k!} = e^{-t\lambda} e^{t\lambda K} = e^{t\lambda(K-I)}$$

This is a simple and elegant representation of the Markov semigroup that makes it easy to understand and analyze distribution dynamics.

For example, if X_0 has distribution ψ, then X_t has distribution

$$\psi P_t = \psi e^{t\lambda(K-I)} \tag{3.12}$$

We just need to plug in λ and K to obtain the entire flow $t \mapsto \psi P_t$.

We will soon extend this representation to the case where S is infinite.

3.7 Distribution Flows for the Inventory Model

Let's apply these ideas to the inventory model described above.

We fix

- the parameters α, b and λ in the inventory model and
- an initial condition $X_0 \sim \psi_0$, where ψ_0 is an arbitrary distribution on S.

The state S is set to $\{0, \dots, b\}$ and the matrix K is defined by (3.11).

Now we run time forward.

We are interested in computing the flow of distributions $t \mapsto \psi_t$, where ψ_t is the distribution of X_t.

According to the theory developed above, we have two options:

Option 1 is to use simulation.

The first step is to simulate many independent observations of the process $(X_t^m)_{m=1}^M$.

(Here m indicates simulation number m, which you might think of as the outcome for firm m.)

Next, for any given t, we define $\hat{\psi}_t \in \mathcal{D}$ as the histogram of observations at time t, or, equivalently the cross-sectional distribution at t:

$$\hat{\psi}_t(x) := \frac{1}{M} \sum_{m=1}^{M} \mathbb{1}\{X_t = x\} \qquad (x \in S)$$

When M is large, $\hat{\psi}_t(x)$ will be close to $\mathbb{P}\{X_t = x\}$ by the law of large numbers.

In other words, in the limit we recover ψ_t.

Option 2 is to insert the parameters into the right hand side of (3.12) and compute ψ_t as $\psi_0 P_t$.

The figure below is created using option 2, with $\alpha = 0.6$, $\lambda = 0.5$ and $b = 10$.

For the initial distribution we pick a binomial distribution.

Since we cannot compute the entire uncountable flow $t \mapsto \psi_t$, we iterate forward 200 steps at time increments $h = 0.1$.

In the figure, hot colors indicate initial conditions and early dates (so that the distribution "cools" over time)
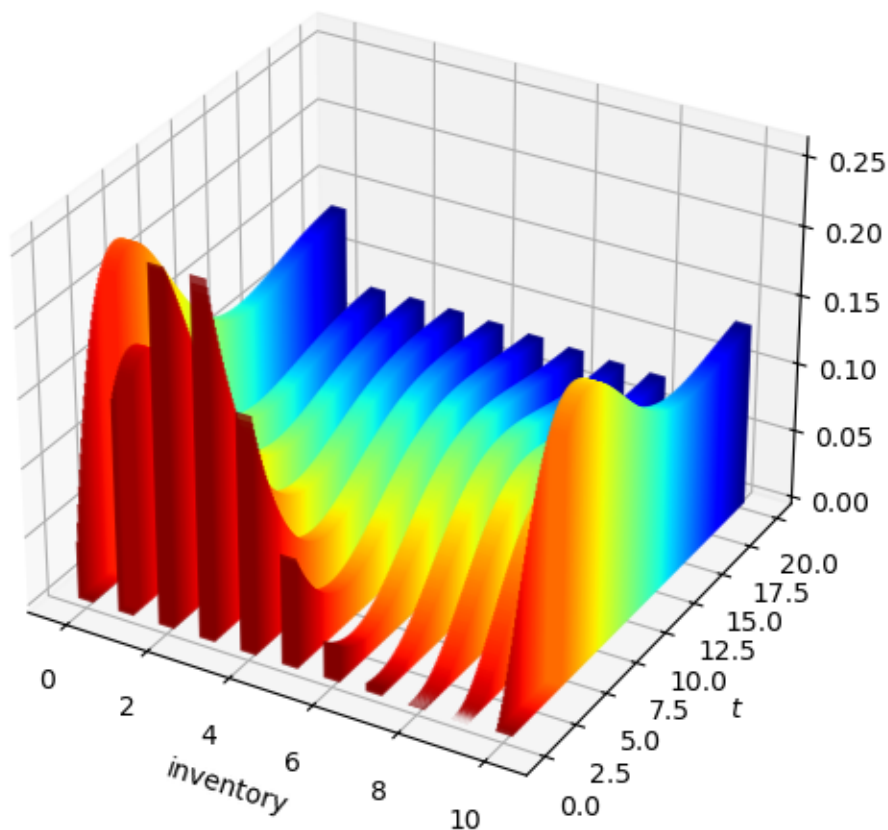
Fig. 3.1: Probability flows for the inventory model.

In the (solved) exercises you will be asked to try to reproduce this figure.

3.8 Exercises

Exercise 3.1

Consider the binary (Bernoulli) distribution where outcomes 0 and 1 each have probability 0.5.

Construct two different random variables with this distribution.

Exercise 3.2

Show by direct calculation that the Poisson matrices (P_t) defined in (3.9) satisfy the semigroup property (3.7).

Hints

- Recall that $P_t(j, k) = 0$ whenever $j > k$.
- Consider using the binomial formula.

Exercise 3.3

Consider the distribution over S^{n+1} previously shown in (3.5), which is

$$\mathbf{P}_\psi^n(x_0, x_1, \dots, x_n) = \psi(x_0)P(x_0, x_1) \times \cdots \times P(x_{n-1}, x_n)$$

Show that, for any Markov chain (X_t) satisfying (3.2) and $X_0 \sim \psi$, the restriction $(X_0, \dots, X_n)$ has joint distribution $\mathbf{P}_\psi^n$.

Exercise 3.4

Try to produce your own version of the figure *Probability flows for the inventory model.*

The initial condition is `ψ_0 = binom.pmf(states, n, 0.25)` where `n = b + 1`.

3.9 Solutions

Note: code is currently not supported in `sphinx-exercise` so code-cell solutions are immediately after this solution block.

Solution to Exercise 3.1

This is easy.

One example is to take U to be uniform on $(0, 1)$ and set $X = 0$ if $U < 0.5$ and 1 otherwise.

Then X has the desired distribution.

Alternatively, we could take Z to be standard normal and set $X = 0$ if $Z < 0$ and 1 otherwise.

Solution to Exercise 3.2

Fixing $s, t \in \mathbb{R}_+$ and $j \leq k$, we have

$$\sum_{i \geq 0} P_s(j, i) P_t(i, k) = e^{-\lambda(s+t)} \sum_{j \leq i \leq k} \frac{(\lambda s)^{i-j}}{(i-j)!} \frac{(\lambda t)^{k-i}}{(k-i)!}$$

$$= e^{-\lambda(s+t)} \lambda^{k-j} \sum_{0 \leq \ell \leq k-j} \frac{s^\ell}{\ell!} \frac{t^{k-j-\ell}}{(k-j-\ell)!}$$

$$= e^{-\lambda(s+t)} \lambda^{k-j} \sum_{0 \leq \ell \leq k-j} \binom{k-j}{\ell} \frac{s^\ell t^{k-j-\ell}}{(k-j)!}$$

Applying the binomial formula, we can write this as

$$\sum_{i \geq 0} P_s(j, i) P_t(i, k) = e^{-\lambda(s+t)} \frac{(\lambda(s+t))^{k-j}}{(k-j)!} = P_{s+t}(j, k)$$

Hence (3.7) holds, and the semigroup property is satisfied.

Solution to Exercise 3.3

Let (X_t) be a Markov chain satisfying (3.2) and $X_0 \sim \psi$.

When $n = 0$, we have $\mathbf{P}_\psi^n = \mathbf{P}_\psi^0 = \psi$, and this agrees with the distribution of the restriction $(X_0, \ldots, X_n) = (X_0)$.

Now suppose the same is true at arbitrary $n-1$, in the sense that the distribution of $(X_0, \ldots, X_{n-1})$ is equal to $\mathbf{P}_\psi^{n-1}$ as defined above.

Then

$$\mathbb{P}\{X_0 = x_0, \ldots, X_n = x_n\} = \mathbb{P}\{X_n = x_n \mid X_0 = x_0, \ldots, X_{n-1} = x_{n-1}\}$$
$$\times \mathbb{P}\{X_0 = x_0, \ldots, X_{n-1} = x_{n-1}\}$$

From the Markov property and the induction hypothesis, the right hand side is

$$P(x_{n-1}, x_n) \mathbf{P}_\psi^{n-1}(x_0, x_1, \ldots, x_{n-1}) = P(x_{n-1}, x_n) \psi(x_0) P(x_0, x_1) \times \cdots \times P(x_{n-2}, x_{n-1})$$

The last expression equals $\mathbf{P}_\psi^n$, which concludes the proof.

Solution to Exercise 3.4

Here is one approach.

(The statements involving `glue` are specific to this book and can be deleted by most readers. They store the output so it can be displayed elsewhere.)

```
α = 0.6
λ = 0.5
b = 10
n = b + 1
states = np.arange(n)
I = np.identity(n)
```

```python
K = np.zeros((n, n))
K[0, -1] = 1
for i in range(1, n):
    for j in range(0, i):
        if j == 0:
            K[i, j] = (1 - α)**(i-1)
        else:
            K[i, j] = α * (1 - α)**(i-j-1)


def P_t(ψ, t):
    return ψ @ expm(t * λ * (K - I))

def plot_distribution_dynamics(ax, ψ_0, steps=200, step_size=0.1):
    ψ = ψ_0
    t = 0.0
    colors = cm.jet_r(np.linspace(0.0, 1, steps))

    for i in range(steps):
        ax.bar(states, ψ, zs=t, zdir='y',
            color=colors[i], alpha=0.8, width=0.4)
        ψ = P_t(ψ, t=step_size)
        t += step_size

    ax.set_xlabel('inventory')
    ax.set_ylabel('$t$')


ψ_0 = binom.pmf(states, n, 0.25)
fig = plt.figure(figsize=(8, 6))
ax = fig.add_subplot(111, projection='3d')
plot_distribution_dynamics(ax, ψ_0)

from myst_nb import glue
glue("flow_fig", fig, display=False)
plt.savefig("_static/lecture_specific/markov_prop/flow_fig.png")

plt.show()
```
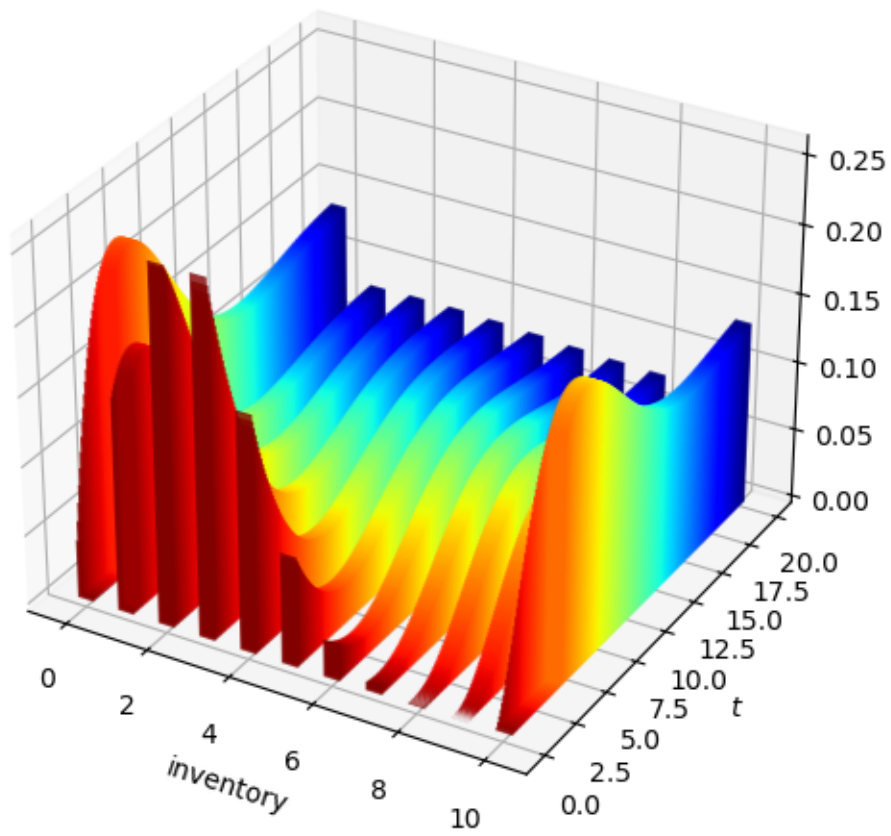
THE KOLMOGOROV BACKWARD EQUATION

In addition to what's in Anaconda, this lecture will need the following libraries:

```
!pip install quantecon
```

4.1 Overview

As models become more complex, deriving analytical representations of the Markov semigroup (P_t) becomes harder.

This is analogous to the idea that solutions to continuous time models often lack analytical solutions.

For example, when studying deterministic paths in continuous time, infinitesimal descriptions (ODEs and PDEs) are often more intuitive and easier to write down than the associated solutions.

(This is one of the shining insights of mathematics, beginning with the work of great scientists such as Isaac Newton.)

We will see in this lecture that the same is true for continuous time Markov chains.

To help us focus on intuition in this lecture, rather than technicalities, the state space is assumed to be finite, with $|S| = n$.

Later we will investigate the case where $|S| = \infty$.

We will use the following imports

```python
import numpy as np
import scipy as sp
import matplotlib.pyplot as plt
import quantecon as qe
from numba import njit

from scipy.linalg import expm
from scipy.stats import binom
```

4.2 State Dependent Jump Intensities

As we have seen, continuous time Markov chains jump between states, and hence can have the form

$$X_t = \sum_{k \geq 0} Y_k \mathbb{1}\{J_k \leq t < J_{k+1}\} \qquad (t \geq 0)$$

where (J_k) are jump times and (Y_k) are the states at each jump.

(We are assuming that $J_k \to \infty$ with probability one, so that X_t is well defined for all $t \geq 0$, but this is always true for when holding times are exponential and the state space is finite.)

In the *previous lecture*,

- the sequence (Y_k) was drawn from a Markov matrix K and called the embedded jump chain, while

- the holding times $W_k := J_k - J_{k-1}$ were IID and $\text{Exp}(\lambda)$ for some constant jump intensity λ.

In this lecture, we will generalize by allowing the jump intensity to vary with the state.

This difference sounds minor but in fact it will allow us to reach full generality in our description of continuous time Markov chains, as clarified below.

4.2.1 Motivation

As a motivating example, recall *the inventory model*, where we assumed that the wait time for the next customer was equal to the wait time for new inventory.

This assumption was made purely for convenience and seems unlikely to hold true.

When we relax it, the jump intensities depend on the state.

4.2.2 Jump Chain Algorithm

We start with three primitives

1. An initial condition ψ,

2. a Markov matrix K on S satisfying $K(x, x) = 0$ for all $x \in S$ and

3. a function λ mapping S to $(0, \infty)$.

The process (X_t)

- starts at state x, draw from ψ,

- waits there for an exponential time W with rate $\lambda(x)$ and then

- updates to a new state y drawn from $K(x, \cdot)$.

Now we take y as the new state for the process and repeat.

Here is the same algorithm written more explicitly:

Algorithm 4.1 (Jump Chain Algorithm)

Inputs $\psi \in \mathcal{D}$, rate function λ, Markov matrix K

Outputs Markov chain (X_t)

1. Draw Y_0 from ψ, set $J_0 = 0$ and $k = 1$.

---

2. Draw W_k independently from $\text{Exp}(\lambda(Y_{k-1}))$.

3. Set $J_k = J_{k-1} + W_k$.

4. Set $X_t = Y_{k-1}$ for t in $[J_{k-1}, J_k)$.

5. Draw Y_k from $K(Y_{k-1}, \cdot)$.

6. Set $k = k + 1$ and go to step 2.

The sequence (W_k) is drawn as an IID sequence and (W_k) and (Y_k) are drawn independently.

The restriction $K(x, x) = 0$ for all x implies that (X_t) actually jumps at each jump time.

4.3 Computing the Semigroup

For the jump process (X_t) with time varying intensities described in the jump chain algorithm, calculating the Markov semigroup is not a trivial exercise.

The approach we adopt is

1. Use probabilistic reasoning to obtain an integral equation that the semigroup must satisfy.

2. Convert the integral equation into a differential equation that is easier to work with.

3. Solve this differential equation to obtain the Markov semigroup (P_t).

The differential equation in question has a special name: the Kolmogorov backward equation.

4.3.1 An Integral Equation

Here is the first step in the sequence listed above.

Lemma 4.1 (An Integral Equation)

The semigroup (P_t) of the jump chain with rate function λ and Markov matrix K obeys the integral equation

$$P_t(x, y) = e^{-t\lambda(x)} I(x, y) + \lambda(x) \int_0^t (KP_{t-\tau})(x, y) e^{-\tau\lambda(x)} d\tau \tag{4.1}$$

for all $t \geq 0$ and x, y in S.

Here (P_t) is the Markov semigroup of (X_t), the process constructed via *Algorithm 4.1*, while $KP_{t-\tau}$ is the matrix product of K and $P_{t-\tau}$.

Proof. Conditioning implicitly on $X_0 = x$, the semigroup (P_t) must satisfy

$$P_t(x, y) = \mathbb{P}\{X_t = y\} = \mathbb{P}\{X_t = y, \ J_1 > t\} + \mathbb{P}\{X_t = y, \ J_1 \leq t\} \tag{4.2}$$

Regarding the first term on the right hand side of (4.2), we have

$$\mathbb{P}\{X_t = y, \ J_1 > t\} = I(x, y) P\{J_1 > t\} = I(x, y) e^{-t\lambda(x)} \tag{4.3}$$

where $I(x, y) = \mathbb{1}\{x = y\}$.

For the second term on the right hand side of (4.2), we have

$$\mathbb{P}\{X_t = y, \ J_1 \le t\} = \mathbb{E}\left[\mathbb{1}\{J_1 \le t\}\mathbb{P}\{X_t = y \mid W_1, Y_1\}\right] = \mathbb{E}\left[\mathbb{1}\{J_1 \le t\}P_{t-J_1}(Y_1, y)\right]$$

Evaluating the expectation and using the independence of J_1 and Y_1, this becomes

$$\mathbb{P}\{X_t = y, \ J_1 \le t\} = \int_0^\infty \mathbb{1}\{\tau \le t\}\sum_z K(x, z)P_{t-\tau}(z, y)\lambda(x)e^{-\tau\lambda(x)}d\tau$$

$$= \lambda(x)\int_0^t \sum_z K(x, z)P_{t-\tau}(z, y)e^{-\tau\lambda(x)}d\tau$$

Combining this result with (4.2) and (4.3) gives (4.1).

4.3.2 Kolmogorov's Differential Equation

We have now confirmed that the semigroup (P_t) associated with the jump chain process (X_t) satisfies (4.1).

Equation (4.1) is important but we can simplify it further without losing information by taking the time derivative.

This leads to our main result for the lecture

Theorem 4.1 (Kolmogorov Backward Equation)

The semigroup (P_t) of the jump chain with rate function λ and Markov matrix K satisfies the **Kolmogorov backward equation**

$$P_t' = QP_t \quad \text{where} \quad Q(x, y) := \lambda(x)(K(x, y) - I(x, y)) \tag{4.4}$$

The derivative on the left hand side of (4.4) is taken element by element, with respect to t, so that

$$P_t'(x, y) = \left(\frac{d}{dt}P_t(x, y)\right) \qquad ((x, y) \in S \times S)$$

The proof that differentiating (4.1) yields (4.4) is an important exercise (see below).

4.3.3 Exponential Solution

The Kolmogorov backward equation is a matrix-valued differential equation.

Recall that, for a scalar differential equation $y_t' = ay_t$ with constant a and initial condition y_0, the solution is $y_t = e^{ta}y_0$.

This, along with $P_0 = I$, encourages us to guess that the solution to Kolmogorov's backward equation (4.4) is

$$P_t = e^{tQ} \tag{4.5}$$

where the right hand side is the matrix exponential, with definition

$$e^{tQ} = \sum_{k\ge 0}\frac{1}{k!}(tQ)^k = I + tQ + \frac{t^2}{2!}Q^2 + \cdots \tag{4.6}$$

Working element by element, it is not difficult to confirm that the derivative of the exponential function $t \mapsto e^{tQ}$ is

$$\frac{d}{dt}e^{tQ} = Qe^{tQ} = e^{tQ}Q \tag{4.7}$$

Hence, differentiating (4.5) gives $P_t' = Qe^{tQ} = QP_t$, which convinces us that the exponential solution satisfies (4.4).

Notice that our solution

$$P_t = e^{tQ} \quad \text{where} \quad Q(x,y) := \lambda(x)(K(x,y) - I(x,y)) \tag{4.8}$$

for the semigroup of the jump process (X_t) associated with the jump matrix K and the jump intensity function $\lambda \colon S \to (0,\infty)$ is consistent with our earlier result.

In particular, we *showed* that, for the model with constant jump intensity λ, we have $P_t = e^{t\lambda(K-I)}$.

This is obviously a special case of (4.8).

4.4 Properties of the Solution

Let's investigate further the properties of the exponential solution.

4.4.1 Checking the Transition Semigroup Properties

While we have confirmed that $P_t = e^{tQ}$ solves the Kolmogorov backward equation, we still need to check that this solution is a Markov semigroup.

Lemma 4.2 (From Jump Chain to Semigroup)

Let λ map S to $\mathbb{R}_+$ and let K be a Markov matrix on S. If $P_t = e^{tQ}$ for all $t \geq 0$, where $Q(x,y) = \lambda(x)(K(x,y) - I(x,y))$, then (P_t) is a Markov semigroup on S.

Proof. Observe first that Q has zero row sums, since

$$\sum_y Q(x,y) = \lambda(x) \sum_y (K(x,y) - I(x,y)) = 0$$

As a small exercise, you can check that, with 1 representing a column vector of ones, the following is true

$$Q \text{ has zero row sums} \iff Q^k 1 = 0 \text{ for all } k \geq 1 \tag{4.9}$$

This implies that $Q^k 1 = 0$ for all k and, as a result, for any $t \geq 0$,

$$P_t 1 = e^{tQ} 1 = I1 + tQ1 + \frac{t^2}{2!}Q^2 1 + \cdots = I1 = 1$$

In other words, each P_t has unit row sums.

Next we check nonnegativity of all elements of P_t (which can easily fail for matrix exponentials).

To this end, adopting an argument from [Stroock, 2013], we set $m := \max_x \lambda(x)$ and $\hat{P} := I + Q/m$.

It is not difficult to check that $\hat{P}$ is a Markov matrix and $Q = m(\hat{P} - I)$.

Recalling that, for matrix exponentials, $e^{A+B} = e^A e^B$ whenever $AB = BA$, we have

$$e^{tQ} = e^{tm(\hat{P}-I)} = e^{-tmI} e^{tm\hat{P}} = e^{-tm}\left(I + tm\hat{P} + \frac{(tm)^2}{2!}\hat{P}^2 + \cdots\right)$$

It is clear from this representation that all entries of e^{tQ} are nonnegative.

Finally, we need to check the continuity condition $P_t(x, y) \to I(x, y)$ as $t \to 0$, which is also part of the definition of a Markov semigroup. This is immediate, in the present case, because the exponential function is continuous, and hence $P_t = e^{tQ} \to e^0 = I$.

We can now be reassured that our solution to the Kolmogorov backward equation is indeed a Markov semigroup.

4.4.2 Uniqueness

Might there be another, entirely different Markov semigroup that also satisfies the Kolmogorov backward equation?

The answer is no: linear ODEs in finite dimensional vector space with constant coefficients and fixed initial conditions (in this case $P_0 = I$) have unique solutions.

In fact it's not hard to supply a proof — see the exercises.

4.5 Application: The Inventory Model

Let us look at a modified version of the inventory model where jump intensities depend on the state.

In particular, the wait time for new inventory will now be exponential at rate γ.

The arrival rate for customers will still be denoted by λ and allowed to differ from γ.

For parameters we take

```
α = 0.6
λ = 0.5
γ = 0.1
b = 10
```

Our plan is to investigate the distribution ψ_T of X_T at $T = 30$.

We will do this by simulating many independent draws of X_T and histogramming them.

(In the exercises you are asked to calculate ψ_T a different way, via (4.8).)

```
@njit
def draw_X(T, X_0, max_iter=5000):
    """
    Generate one draw of X_T given X_0.
    """

    J, Y = 0, X_0
    m = 0

    while m < max_iter:
        s = 1/γ if Y == 0 else 1/λ
        W = np.random.exponential(scale=s)   # W ~ E(λ)
        J += W
        if J >= T:
            return Y
        # Otherwise update Y
        if Y == 0:
            Y = b
        else:
```

(continues on next page)

```
            U = np.random.geometric(α)
            Y = Y - min(Y, U)
        m += 1


@njit
def independent_draws(T=10, num_draws=100):
    "Generate a vector of independent draws of X_T."

    draws = np.empty(num_draws, dtype=np.int64)

    for i in range(num_draws):
        X_0 = np.random.binomial(b+1, 0.25)
        draws[i] = draw_X(T, X_0)

    return draws
```

```
T = 30
n = b + 1
draws = independent_draws(T, num_draws=100_000)
fig, ax = plt.subplots()

ax.bar(range(n), [np.mean(draws == i) for i in range(n)], width=0.8, alpha=0.6)
ax.set_xlabel("inventory", fontsize=14)

plt.show()
```
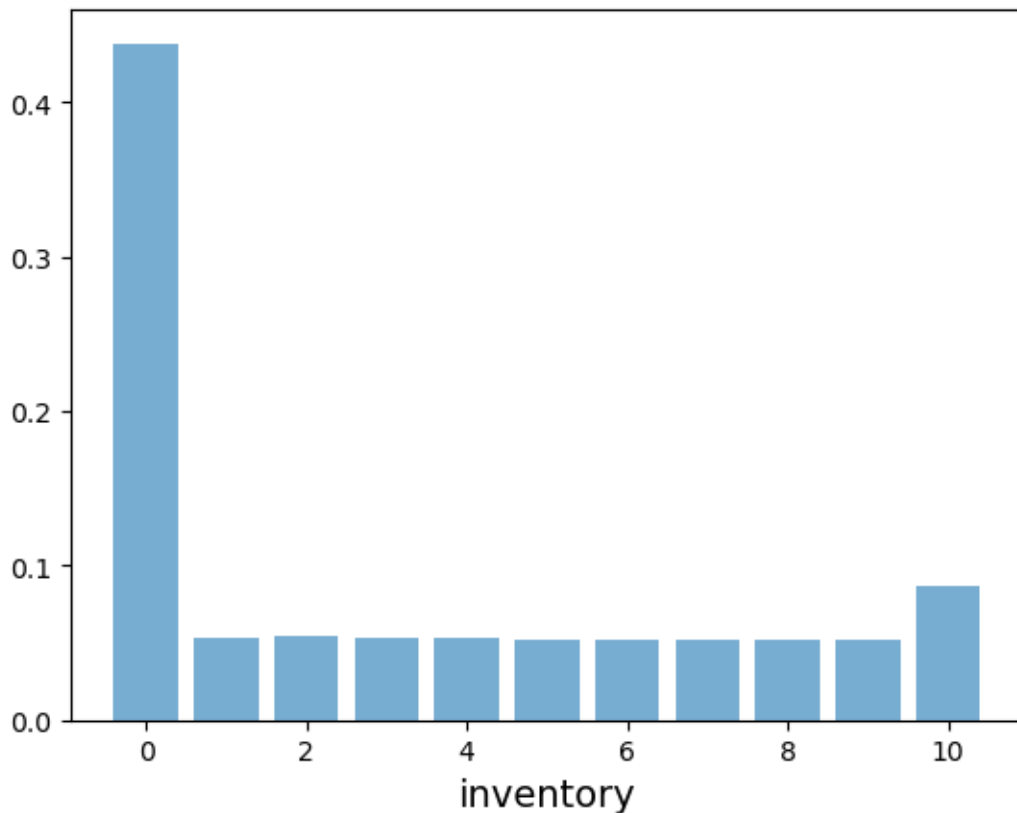
If you experiment with the code above, you will see that the large amount of mass on zero is due to the low arrival rate γ for inventory.

4.6 Exercises

Exercise 4.1

In the discussion above, we generated an approximation of ψ_T when $T = 30$, the initial condition is Binomial$(n, 0.25)$ and parameters are set to

```
α = 0.6
λ = 0.5
γ = 0.1
b = 10
```

The calculation was done by simulating independent draws and histogramming.

Try to generate the same figure using (4.8) instead, modifying code from *our lecture* on the Markov property.

Exercise 4.2

Prove that differentiating (4.1) at each (x, y) yields (4.4).

Exercise 4.3

We claimed above that the solution $P_t = e^{tQ}$ is the unique Markov semigroup satisfying the backward equation $P'_t = QP_t$.

Try to supply a proof.

(This is not an easy exercise but worth thinking about in any case.)

4.7 Solutions

Note: code is currently not supported in `sphinx-exercise` so code-cell solutions are immediately after this solution block.

Solution to Exercise 4.1

Here is one solution:

```
α = 0.6
λ = 0.5
γ = 0.1
b = 10
```
(continues on next page)

Chapter 4. The Kolmogorov Backward Equation

```python
states = np.arange(n)
I = np.identity(n)

# Embedded jump chain matrix
K = np.zeros((n, n))
K[0, -1] = 1
for i in range(1, n):
    for j in range(0, i):
        if j == 0:
            K[i, j] = (1 - α)**(i-1)
        else:
            K[i, j] = α * (1 - α)**(i-j-1)

# Jump intensities as a function of the state
r = np.ones(n) * λ
r[0] = γ

# Q matrix
Q = np.empty_like(K)
for i in range(n):
    for j in range(n):
        Q[i, j] = r[i] * (K[i, j] - I[i, j])

def P_t(ψ, t):
    return ψ @ expm(t * Q)

ψ_0 = binom.pmf(states, n, 0.25)
ψ_T = P_t(ψ_0, T)

fig, ax = plt.subplots()

ax.bar(range(n), ψ_T, width=0.8, alpha=0.6)
ax.set_xlabel("inventory", fontsize=14)

plt.show()
```
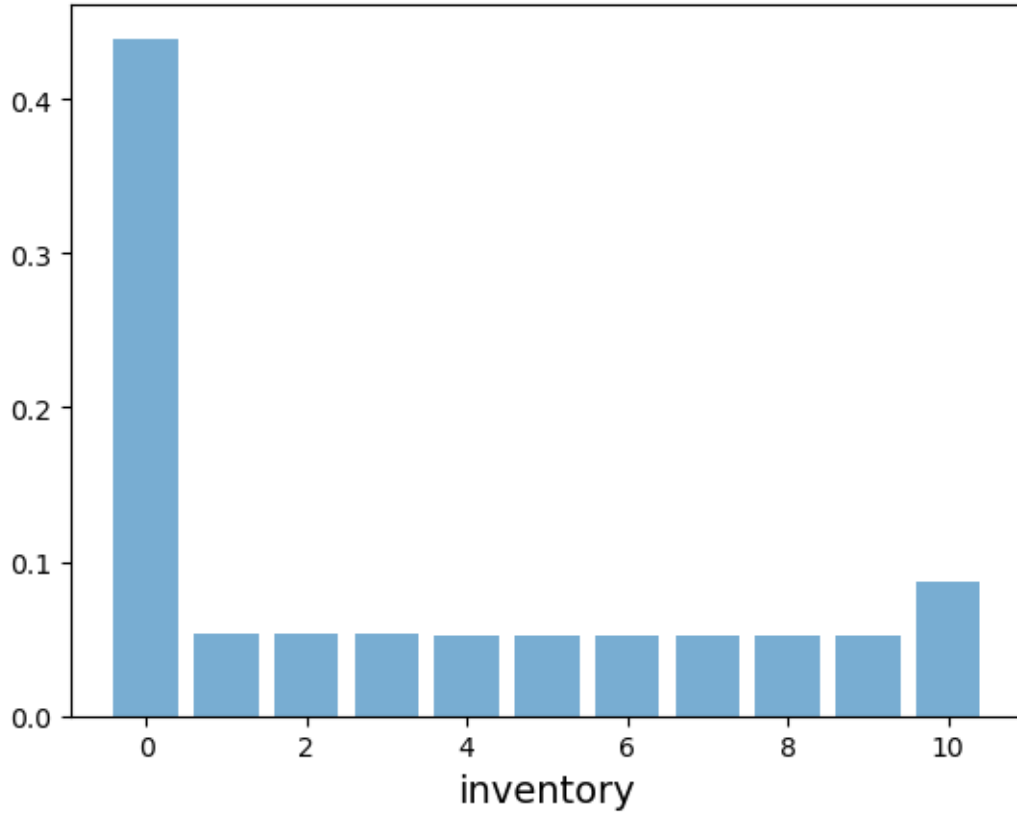
Solution to Exercise 4.2

One can easily verify that, when f is a differentiable function and $\alpha > 0$, we have

$$g(t) = e^{-t\alpha} f(t) \quad \implies \quad g'(t) = e^{-t\alpha} f'(t) - \alpha g(t) \tag{4.10}$$

Note also that, with the change of variable $s = t - \tau$, we can rewrite (4.1) as

$$P_t(x, y) = e^{-t\lambda(x)} \left\{ I(x, y) + \lambda(x) \int_0^t (KP_s)(x, y) e^{s\lambda(x)} ds \right\} \tag{4.11}$$

Applying (4.10) yields

$$P'_t(x, y) = e^{-t\lambda(x)} \left\{ \lambda(x)(KP_t)(x, y) e^{t\lambda(x)} \right\} - \lambda(x) P_t(x, y)$$

After minor rearrangements this becomes

$$P'_t(x, y) = \lambda(x)[(K - I)P_t](x, y)$$

which is identical to (4.4).

Solution to Exercise 4.3

Here is one proof of uniqueness.

Suppose that $(\hat{P}_t)$ is another Markov semigroup satisfying $P'_t = QP_t$.

Fix $t > 0$ and let V_s be defined by $V_s = P_s \hat{P}_{t-s}$ for all $s \geq 0$.

Note that $V_0 = \hat{P}_t$ and $V_t = P_t$.

Note also that $s \mapsto V_s$ is differentiable, with derivative

$$V_s' = P_s' \hat{P}_{t-s} - P_s \hat{P}_{t-s}' = P_s Q \hat{P}_{t-s} - P_s Q \hat{P}_{t-s} = 0$$

where, in the second last equality, we used (4.7).

Hence V_s is constant, so our previous observations $V_0 = \hat{P}_t$ and $V_t = P_t$ now yield $\hat{P}_t = P_t$.

Since t was arbitrary, the proof is now done.

THE KOLMOGOROV FORWARD EQUATION

In addition to what's in Anaconda, this lecture will need the following libraries:

```
!pip install quantecon
```

5.1 Overview

In this lecture we approach continuous time Markov chains from a more analytical perspective.

The emphasis will be on describing distribution flows through vector-valued differential equations and their solutions.

These distribution flows show how the time t distribution associated with a given Markov chain (X_t) changes over time.

Distribution flows will be identified with initial value problems generated by autonomous linear ordinary differential equations (ODEs) in vector space.

We will see that the solutions of these flows are described by Markov semigroups.

This leads us back to the theory we have already constructed – some care will be taken to clarify all the connections.

In order to avoid being distracted by technicalities, we continue to defer our treatment of infinite state spaces, assuming throughout this lecture that $|S| = n$.

As before, $\mathcal{D}$ is the set of all distributions on S.

We will use the following imports

```python
import numpy as np
import scipy as sp
import matplotlib.pyplot as plt
import quantecon as qe
from numba import njit
from scipy.linalg import expm

from matplotlib import cm
from mpl_toolkits.mplot3d import Axes3D
from mpl_toolkits.mplot3d.art3d import Poly3DCollection
```

5.2 From Difference Equations to ODEs

Previously we generated this figure, which shows how distributions evolve over time for the inventory model under a certain parameterization:
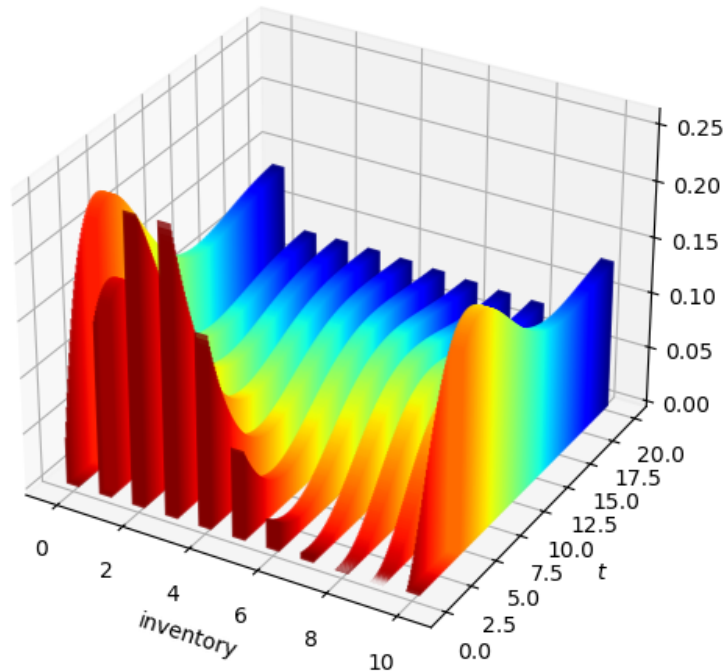
Fig. 5.1: Probability flows for the inventory model.

(Hot colors indicate early dates and cool colors denote later dates.)

We also learned how this flow is related to the Kolmogorov backward equation, which is an ODE.

In this section we examine distribution flows and their connection to ODEs and continuous time Markov chains more systematically.

5.2.1 Review of the Discrete Time Case

Let (X_t) be a discrete time Markov chain with Markov matrix P.

Recall that, in the discrete time case, the distribution ψ_t of X_t updates according to

$$\psi_{t+1} = \psi_t P, \qquad \psi_0 \text{ a given element of } \mathcal{D},$$

where distributions are understood as row vectors.

Here's a visualization for the case $S = \{0, 1, 2\}$, so that $\mathcal{D}$ is the standard simplex in $\mathbb{R}^3$.

The initial condition is `(0, 0, 1)` and the Markov matrix is

```
P  =  ((0.9,  0.1,  0.0),
       (0.4,  0.4,  0.2),
       (0.1,  0.1,  0.8))
```
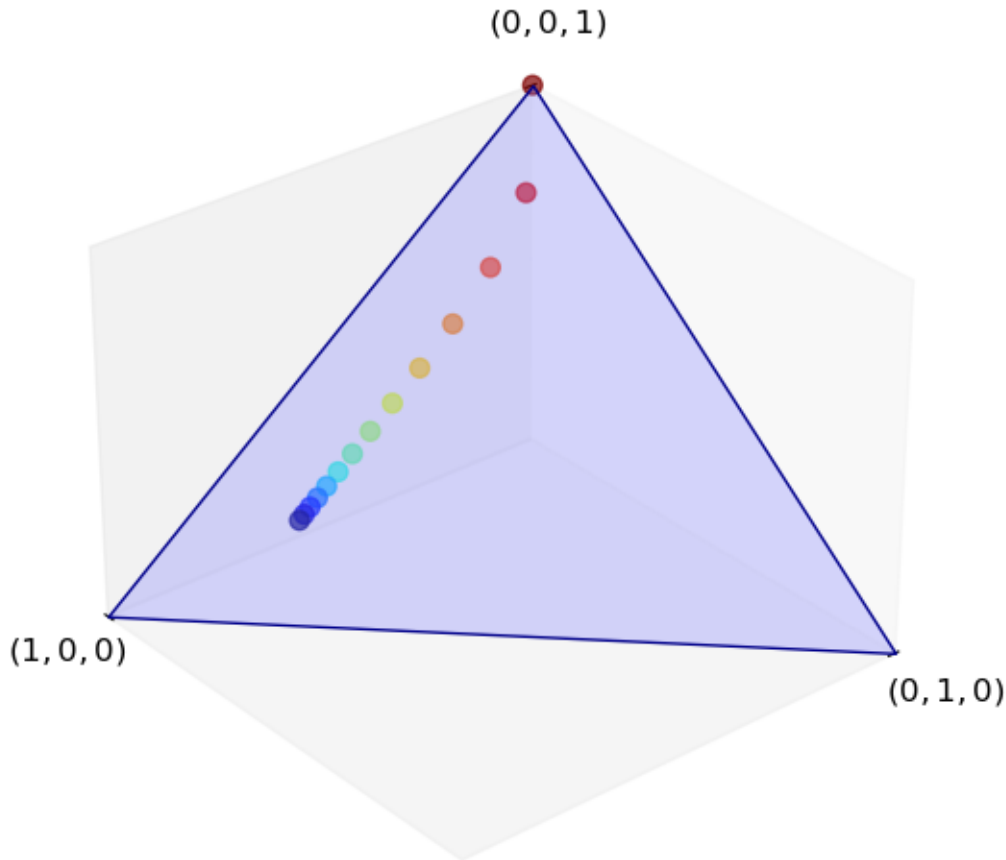

There's a sense in which a discrete time Markov chain "is" a homogeneous linear difference equation in distribution space.

To clarify this, suppose we take G to be a linear map from $\mathcal{D}$ to itself and write down the difference equation

$$\psi_{t+1} = G(\psi_t) \quad \text{with } \psi_0 \in \mathcal{D} \text{ given.} \tag{5.1}$$

Because G is a linear map from a finite dimensional space to itself, it can be represented by a matrix.

Moreover, a matrix P is a Markov matrix if and only if $\psi \mapsto \psi P$ sends $\mathcal{D}$ into itself (check it if you haven't already).

So, under the stated conditions, our difference equation (5.1) uniquely identifies a Markov matrix, along with an initial condition ψ_0.

Together, these objects identify the joint distribution of a discrete time Markov chain, as *previously described*.

5.2.2 Shifting to Continuous Time

We have just argued that a discrete time Markov chain can be identified with a linear difference equation evolving in $\mathcal{D}$.

This strongly suggests that a continuous time Markov chain can be identified with a linear ODE evolving in $\mathcal{D}$.

This intuition is correct and important.

The rest of the lecture maps out the main ideas.

5.3 ODEs in Distribution Space

Consider linear differential equation given by

$$\psi_t' = \psi_t Q, \qquad \psi_0 \text{ a given element of } \mathcal{D}, \tag{5.2}$$

where

- Q is an $n \times n$ matrix,
- distributions are again understood as row vectors, and
- derivatives are taken element by element, so that

$$\psi_t' = \begin{pmatrix} \frac{d}{dt}\psi_t(x_1) & \cdots & \frac{d}{dt}\psi_t(x_n) \end{pmatrix}$$

5.3.1 Solutions to Linear Vector ODEs

Using the matrix exponential, the unique solution to the initial value problem (5.2) is

$$\psi_t = \psi_0 P_t \quad \text{where } P_t := e^{tQ} \tag{5.3}$$

To check that (5.3) is a solution, we use (4.7) again to get

$$\frac{d}{dt}P_t = Qe^{tQ} = e^{tQ}Q$$

The first equality can be written as $P_t' = QP_t$ and this is just the *Kolmogorov backward equation*.

The second equality can be written as

$$P_t' = P_t Q$$

and is called the **Kolmogorov forward equation**.

Applying the Kolmogorov forward equation, we obtain

$$\frac{d}{dt}\psi_t = \frac{d}{dt}\psi_0 P_t = \psi_0 \frac{d}{dt}P_t = \psi_0 P_t Q = \psi_t Q$$

This confirms that (5.3) solves (5.2).

Here's an example of three distribution flows with dynamics generated by (5.2), one starting from each vertex.

The code uses (5.3) with matrix Q given by

```
Q = ((-3,  2,   1),
     (3,  -5,   2),
     (4,   6, -10))
```
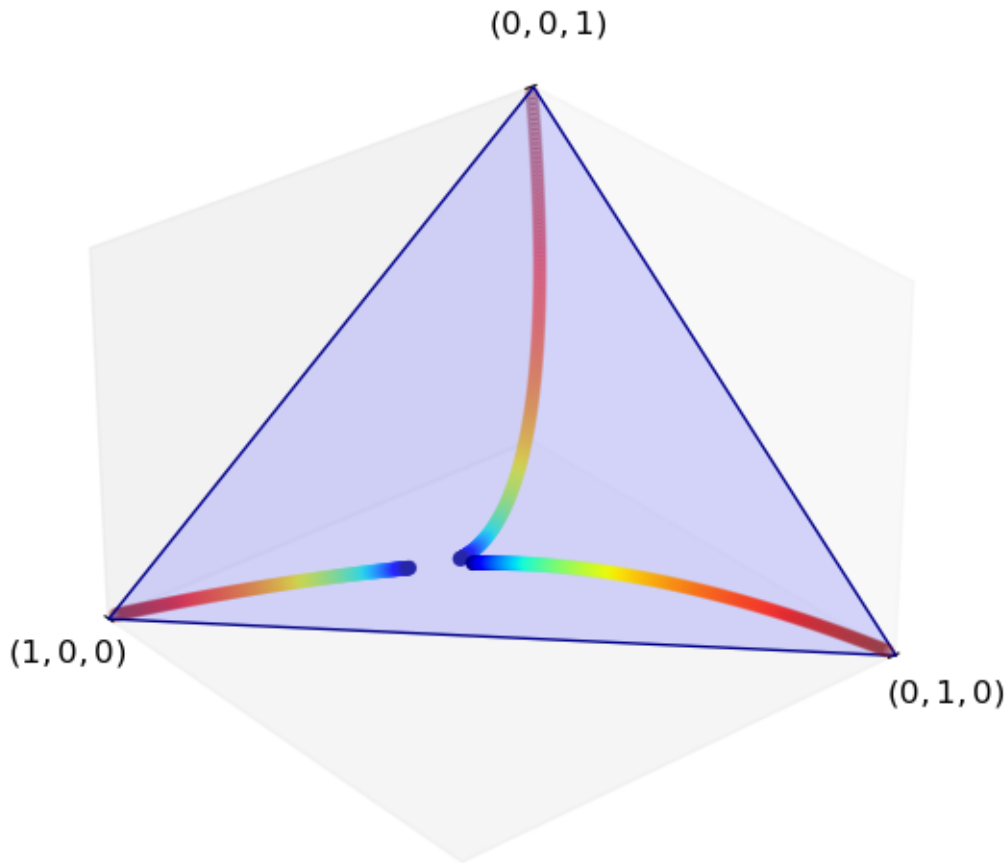

(Distributions cool over time, so initial conditions are hot colors.)

5.3.2 Forwards vs Backwards Equations

As the above discussion shows, we can take the Kolmogorov forward equation $P_t' = P_t Q$ and premultiply by any distribution ψ_0 to get the distribution ODE $\psi_t' = \psi_t Q$.

In this sense, we can understand the Kolmogorov forward equation as pushing distributions forward in time.

Analogously, we can take the Kolmogorov backward equation $P_t' = Q P_t$ and postmultiply by any vector h to get

$$(P_t h)' = Q P_t h$$

Recalling that $(P_t h)(x) = \mathbb{E}\left[h(X_t) \mid X_0 = x\right]$, this vector ODE tells us how expectations evolve, conditioning backward to time zero.

Both the forward and the backward equations uniquely pin down the same solution $P_t = e^{tQ}$ when combined with the initial condition $P_0 = I$.

5.3.3 Matrix- vs Vector-Valued ODEs

The ODE $\psi'_t = \psi_t Q$ is sometimes called the **Fokker–Planck equation** (although this terminology is most commonly used in the context of diffusions).

It is a vector-valued ODE that describes the evolution of a particular distribution path.

By comparison, the Kolmogorov forward equation is (like the backward equation) a differential equation in matrices.

(And matrices are really maps, which send vectors into vectors.)

Operating at this level is less intuitive and more abstract than working with the Fokker–Planck equation.

But, in the end, the object that we want to describe is a Markov semigroup.

The Kolmogorov forward and backward equations are the ODEs that define this fundamental object.

5.3.4 Preserving Distributions

In the simulation above, Q was chosen with some care, so that the flow remains in $\mathcal{D}$.

What are the exact properties we require on Q such that ψ_t is always in $\mathcal{D}$?

This is an important question, because we are setting up an exact correspondence between linear ODEs that evolve in $\mathcal{D}$ and continuous time Markov chains.

Recall that the linear update rule $\psi \mapsto \psi P$ is invariant on $\mathcal{D}$ if and only if P is a Markov matrix.

So now we can rephrase our key question regarding invariance on $\mathcal{D}$:

What properties do we need to impose on Q so that $P_t = e^{tQ}$ is a Markov matrix for all t?

A square matrix Q is called an **intensity matrix** if Q has zero row sums and $Q(x, y) \geq 0$ whenever $x \neq y$.

Theorem 5.1

If Q is a matrix on S and $P_t := e^{tQ}$, then the following statements are equivalent:

1. P_t is a Markov matrix for all t.
2. Q is an intensity matrix.

The proof is related to that of *Lemma 4.2* and is found as a solved exercise below.

Corollary 5.1

If Q is an intensity matrix on finite S and $P_t = e^{tQ}$ for all $t \geq 0$, then (P_t) is a Markov semigroup.

We call (P_t) the Markov semigroup **generated** by Q.

Later we will see that this result extends to the case $|S| = \infty$ under some mild restrictions on Q.

5.4 Jump Chains

Let's return to the chain (X_t) created from jump chain pair (λ, K) in *Algorithm 4.1*.

We found that the semigroup is given by

$$P_t = e^{tQ} \quad \text{where} \quad Q(x, y) := \lambda(x)(K(x, y) - I(x, y))$$

Using the fact that K is a Markov matrix and the jump rate function λ is nonnegative, you can easily check that Q satisfies the definition of an intensity matrix.

Hence (P_t), the Markov semigroup for the jump chain (X_t), is the semigroup generated by the intensity matrix $Q(x, y) = \lambda(x)(K(x, y) - I(x, y))$.

We can differentiate $P_t = e^{tQ}$ to obtain the Kolmogorov forward equation $P_t' = P_t Q$.

We can then premultiply by $\psi_0 \in \mathcal{D}$ to get $\psi_t' = \psi_t Q$, which is the Fokker–Planck equation.

More explicitly, for given $y \in S$,

$$\psi_t'(y) = \sum_{x \neq y} \psi_t(x) \lambda(x) K(x, y) - \psi_t(y) \lambda(y)$$

The rate of probability flow into y is equal to the inflow from other states minus the outflow.

5.5 Summary

We have seen that any intensity matrix Q on S defines a Markov semigroup via $P_t = e^{tQ}$.

Henceforth, we will say that (X_t) is **a Markov chain with intensity matrix** Q if (X_t) is a Markov chain with Markov semigroup (e^{tQ}).

While our discussion has been in the context of a finite state space, later we will see that these ideas carry over to an infinite state setting under mild restrictions.

We have also hinted at the fact that *every* continuous time Markov chain is a Markov chain with intensity matrix Q for some suitably chosen Q.

Later we will prove this to be universally true when S is finite and true under mild conditions when S is countably infinite.

Intensity matrices are important because

1. they are the natural infinitesimal descriptions of Markov semigroups,

2. they are often easy to write down in applications and

3. they provide an intuitive description of dynamics.

Later, we will see that, for a given intensity matrix Q, the elements are understood as follows:

- when $x \neq y$, the value $Q(x, y)$ is the "rate of leaving x for y" and

- $-Q(x, x) \geq 0$ is the "rate of leaving x" .

5.6 Exercises

Exercise 5.1

Let (P_t) be a Markov semigroup such that $t \mapsto P_t(x, y)$ is differentiable at all $t \geq 0$ and $(x, y) \in S \times S$.

(The derivative at $t = 0$ is the usual right derivative.)

Define (pointwise, at each (x, y)),

$$Q := P_0' = \lim_{h \downarrow 0} \frac{P_h - I}{h} \qquad (5.4)$$

Assuming that this limit exists, and hence Q is well-defined, show that

$$P_t' = P_t Q \quad \text{and} \quad P_t' = Q P_t$$

both hold. (These are the Kolmogorov forward and backward equations.)

Exercise 5.2

Recall *our model* of jump chains with state-dependent jump intensities given by rate function $x \mapsto \lambda(x)$.

After a wait time with exponential rate $\lambda(x) \in (0, \infty)$, the state transitions from x to y with probability $K(x, y)$.

We found that the associated semigroup (P_t) satisfies the Kolmogorov backward equation $P_t' = Q P_t$ with

$$Q(x, y) := \lambda(x)(K(x, y) - I(x, y)) \qquad (5.5)$$

Show that Q is an intensity matrix and that (5.4) holds.

Exercise 5.3

Prove *Theorem 5.1* by adapting the arguments in *Lemma 4.2*. (This is nontrivial but worth at least trying.)

Hint: The constant m in the proof can be set to $\max_x |Q(x, x)|$.

5.7 Solutions

Solution to Exercise 5.1

Let (P_t) be a Markov semigroup and let Q be as defined in the statement of the exercise.

Fix $t \geq 0$ and $h > 0$.

Combining the semigroup property and linearity with the restriction $P_0 = I$, we get

$$\frac{P_{t+h} - P_t}{h} = \frac{P_t P_h - P_t}{h} = \frac{P_t(P_h - I)}{h}$$

Taking $h \downarrow 0$ and using the definition of Q give $P_t' = P_t Q$, which is the Kolmogorov forward equation.

For the backward equation we observe that

$$\frac{P_{t+h} - P_t}{h} = \frac{P_h P_t - P_t}{h} = \frac{(P_h - I)P_t}{h}$$

also holds. Taking $h \downarrow 0$ gives the Kolmogorov backward equation.

Solution to Exercise 5.2

Let Q be as defined in (5.5).

We need to show that Q is nonnegative off the diagonal and has zero row sums.

The first assertion is immediate from nonnegativity of K and λ.

For the second, we use the fact that K is a Markov matrix, so that, with 1 as a column vector of ones,

$$Q1 = \lambda(K1 - 1) = \lambda(1 - 1) = 0$$

Solution to Exercise 5.3

Suppose that Q is an intensity matrix, fix $t \geq 0$ and set $P_t = e^{tQ}$.

The proof from *Lemma 4.2* that P_t has unit row sums applies directly to the current case.

The proof of nonnegativity of P_t can be applied after some modifications.

To this end, set $m := \max_x |Q(x, x)|$ and $\hat{P} := I + Q/m$.

You can check that $\hat{P}$ is a Markov matrix and that $Q = m(\hat{P} - I)$.

The rest of the proof of nonnegativity of P_t is unchanged and we will not repeat it.

We conclude that P_t is a Markov matrix.

Regarding the converse implication, suppose that $P_t = e^{tQ}$ is a Markov matrix for all t and let 1 be a column vector of ones.

Because P_t has unit row sums and differentiation is linear, we can employ the Kolmogorov backward equation to obtain

$$Q1 = QP_t 1 = \left(\frac{d}{dt}P_t\right)1 = \frac{d}{dt}(P_t 1) = \frac{d}{dt}1 = 0$$

Hence Q has zero row sums.

We can use the definition of the matrix exponential to obtain, for any x, y and $t \geq 0$,

$$P_t(x, y) = \mathbb{1}\{x = y\} + tQ(x, y) + o(t) \tag{5.6}$$

From this equality and the assumption that P_t is a Markov matrix for all t, we see that the off diagonal elements of Q must be nonnegative.

Hence Q is an intensity matrix.

SEMIGROUPS AND GENERATORS

6.1 Overview

We have seen in previous lectures that every intensity matrix generates a Markov semigroup.

We have also hinted that the pairing is one-to-one, in a sense to be made precise.

To clarify these ideas, we start in an abstract setting, with an arbitrary initial value problem.

In this setting we introduce general operator semigroups and their generators.

Once this is done, we will be able to return to the Markov case and fully clarify the connection between intensity matrices and Markov semigroups.

The material below is relatively technical, with most of the complications driven by the fact that the state space can be infinite.

Such technicalities are hard to avoid, since so many interesting Markov chains do have infinite state spaces.

- Our very first example – the Poisson process – has an infinite state space.

- Another example is the study of queues, which often have no natural upper bound.[1]

Readers are assumed to have some basic familiarity with Banach spaces.

6.2 Motivation

The general theory of continuous semigroups of operators is motivated by the problem of solving linear ODEs in infinite dimensional spaces.[2]

More specifically, the challenge is to solve initial value problems such as

$$x_t' = Ax_t, \quad x_0 \text{ given} \tag{6.1}$$

where

- x_t takes value in a Banach space at each time t,

- A is a linear operator and

- the time derivative x_t' uses a definition appropriate for a Banach space.

---

[1] In fact a major concern with queues is that their length does not explode. This issue cannot be properly explored unless the state space is allowed to be infinite.

[2] An excellent introduction to operator semigroups, combined with applications to PDEs and Markov processes, can be found in [Applebaum, 2019].

This problem is also called the "abstract Cauchy problem".

Why do we need to solve such problems?

One example comes from PDEs.

PDEs tell us how functions change over time, starting from an infinitesimal description.

When x_t is a point in a function space, this fits into the framework of (6.1).

Another example comes from Markov processes, where, as we have seen, the flow of distributions over time can be represented as a linear ODE in distribution space.

If the number of state is infinite, then the space of distributions is infinite dimensional.

This is another version of (6.1), and we return to it after a discussion of the general theory.

To give a high level view of the results below, the solution to the Cauchy problem is represented as a trajectory $t \mapsto U_t x_0$ from the initial value x_0 under a semigroup of maps (U_t).

The operator A in (6.1) is called the "generator" of (U_t) and is its infinitesimal description.

6.3 Preliminaries

Throughout this lecture, $(\mathbb{B}, \|\cdot\|)$ is a Banach space.

6.3.1 The Space of Linear Operators

You will recall that a **linear operator** on $\mathbb{B}$ is a map A from $\mathbb{B}$ to itself satisfying

$$A(\alpha g + \beta h) = \alpha Ag + \beta Ah, \quad \forall\, g, h \in \mathbb{B}, \ \alpha, \beta \in \mathbb{R}$$

The operator A is called **bounded** if

$$\|A\| := \sup_{g \in \mathbb{B},\, \|g\| \leq 1} \|Ag\| < \infty \tag{6.2}$$

This is the usual definition of a bounded linear operator on a normed linear space.

The set of all bounded linear operators on $\mathbb{B}$ is denoted by $\mathcal{L}(\mathbb{B})$ and is itself a Banach space.

Sums and scalar products of elements of $\mathcal{L}(\mathbb{B})$ are defined in the usual way, so that, for $\alpha \in \mathbb{R}$, $A, B \in \mathcal{L}(\mathbb{B})$ and $g \in \mathbb{B}$, we have

$$(A + B)g = Ag + Bg, \quad (\alpha A)g = \alpha(Ag)$$

and so on.

We write AB to indicate composition of the operators $A, B \in \mathcal{L}(\mathbb{B})$.

The value defined in (6.2) is called the **operator norm** of A and, as suggested by the notation, is a norm on $\mathcal{L}(\mathbb{B})$.

In addition to being a norm, it enjoys the submultiplicative property $\|AB\| \leq \|A\|\|B\|$ for all $A, B \in \mathcal{L}(\mathbb{B})$.

Let I be the identity in $\mathcal{L}(\mathbb{B})$, satisfying $Ig = g$ for all $g \in \mathbb{B}$.

(In fact $\mathcal{L}(\mathbb{B})$ is a unital Banach algebra when multiplication is identified with operator composition and I is adopted as the unit.)

6.3.2 The Exponential Function

Given $A \in \mathcal{L}(\mathbb{B})$, the exponential of A is the element of $\mathcal{L}(\mathbb{B})$ defined as

$$e^A = \sum_{k \geq 0} \frac{A^k}{k!} = I + A + \frac{A^2}{2!} + \cdots \tag{6.3}$$

This is the same as the definition for the matrix exponential. The exponential function arises naturally as the solution to ODEs in Banach space, one example of which (as we shall see) is distribution flows associated with continuous time Markov chains.

The exponential map has the following properties:

- For each $A \in \mathcal{L}(\mathbb{B})$, the operator e^A is a well defined element of $\mathcal{L}(\mathbb{B})$ with $\|e^A\| \leq e^{\|A\|}$.[3]

- $e^0 = I$, where 0 is the zero element of $\mathcal{L}(\mathbb{B})$.

- If $A, B \in \mathcal{L}(\mathbb{B})$ and $AB = BA$, then $e^{A+B} = e^A e^B$

- If $A \in \mathcal{L}(\mathbb{B})$, then e^A is invertible and $(e^A)^{-1} = e^{-A}$.

The last fact is easily checked from the previous ones.

6.3.3 Operator Calculus

Consider a function

$$\mathbb{R}_+ \ni t \mapsto U_t \in \mathcal{L}(\mathbb{B})$$

which we can think of as a time path in $\mathcal{L}(\mathbb{B})$, such as a flow of Markov operators.

We say that this function is **differentiable at** $\tau \in \mathbb{R}_+$ if there exists an element T of $\mathcal{L}(\mathbb{B})$ such that

$$\frac{U_{\tau+h} - U_\tau}{h} \to T \quad \text{as } h \to 0 \tag{6.4}$$

In this case, T is called the **derivative** of the function $t \mapsto U_t$ at τ and we write

$$T = U'_\tau \quad \text{or} \quad T = \frac{d}{dt} U_t \Big|_{t=\tau}$$

(Convergence of operators is in operator norm. If $\tau = 0$, then the limit $h \to 0$ in (6.4) is the right limit.)

Example

If $U_t = tV$ for some fixed $V \in \mathcal{L}(\mathbb{B})$, then it is easy to see that V is the derivative of $t \mapsto U_t$ at every $t \in \mathbb{R}_+$.

Example

In *our discussion* of the Kolmogorov forward equation when S is finite, we introduced the derivative of a map $t \mapsto P_t$, where each P_t is a matrix on S.

The derivative was defined by differentiating P_t element-by-element.

This coincides with the operator-theoretic definition in (6.4) when S is finite, because then the space $\mathcal{L}(\ell_1)$, which consists of all bounded linear operators on ℓ_1, is finite dimensional, and hence pointwise and norm convergence coincide.

---

[3] Convergence of the sum in (6.3) follows from boundedness of A and the fact that $\mathcal{L}(\mathbb{B})$ is a Banach space.

Analogous to the matrix and scalar cases, we have the following result:

Lemma (Differentiability of the Exponential Curve)

For all $A \in \mathcal{L}(\mathbb{B})$, the exponential curve $t \mapsto e^{tA}$ is everywhere differentiable and

$$\frac{d}{dt}e^{tA} = e^{tA}A = Ae^{tA} \tag{6.5}$$

The proof is a (solved) exercise (see below).

6.4 Semigroups and Generators

For continuous time Markov chains where the state space S is finite, we saw that Markov semigroups often take the form $P_t = e^{tQ}$ for some intensity matrix Q.

This is ideal because the entire semigroup is characterized in a simple way by its infinitesimal description Q.

It turns out that, when S is finite, this is always true: if (P_t) is a Markov semigroup, then there exists an intensity matrix Q satisfying $P_t = e^{tQ}$ for all t.

Moreover, this statement is again true when S is infinite, provided that some restrictions are placed on the semigroup.

Our aim is to make these statements precise, starting in an abstract setting and then specializing.

6.4.1 Operator Semigroups

Let U_t be an element of $\mathcal{L}(\mathbb{B})$ for all $t \in \mathbb{R}_+$

We say that (U_t) is an **evolution semigroup** on $\mathcal{L}(\mathbb{B})$ if $U_0 = I$ and $U_{s+t} = U_s U_t$ for all $s, t \geq 0$.

The idea is that (U_t) generates a path in $\mathbb{B}$ from any starting point $g \in \mathbb{B}$, so that $U_t g$ is interpreted as the location of the state after t units of time.

An evolution semigroup (U_t) is called

- a C_0 **semigroup** on $\mathbb{B}$ if, for each $g \in \mathbb{B}$, the map $t \mapsto U_t g$ from $\mathbb{R}_+$ to $\mathbb{B}$ is continuous, and

- a **uniformly continuous semigroup** on $\mathbb{B}$ if the map $t \mapsto U_t$ from $\mathbb{R}_+$ to $\mathcal{L}(\mathbb{B})$ is continuous.

In what follows we abbreviate "uniformly continuous" to UC.[4]

Example (Exponential curves are UC semigroups)

If $U_t = e^{tA}$ for $t \in \mathbb{R}_+$ and $A \in \mathcal{L}(\mathbb{B})$, then (U_t) is a uniformly continuous semigroup on $\mathbb{B}$.

The claim that (U_t) is an evolution semigroup follows directly from the properties of the exponential function given above.

Uniform continuity can be established using arguments similar to those in the proof of differentiability in *Lemma 6.1*.

Since norm convergence on $\mathcal{L}(\mathbb{B})$ implies pointwise convergence, every uniformly continuous semigroup is a C_0 semigroup.

The reverse is certainly not true — there are many important C_0 semigroups that fail to be uniformly continuous.

---

[4] Be careful: the definition of a UC semigroup requires that $t \mapsto U_t$ is continuous as a map into $\mathcal{L}(\mathbb{B})$, rather than uniformly continuous. The UC terminology comes about because, for a UC semigroup, we have, by definition of the operator norm, $\sup_{\|g\| \leq 1} \|U_s g - U_t g\| \to 0$ when $s \to t$.

In fact semigroups associated with PDEs, diffusions and other Markov processes on continuous state spaces are typically C_0 but not uniformly continuous.

There are also important examples of Markov semigroups on infinite discrete state spaces that fail to be uniformly continuous.

However, we will soon see that, for most continuous time Markov chains used in applications, the semigroups are uniformly continuous.

6.4.2 Generators

Consider a continuous time Markov chain on a finite state space with intensity matrix Q.

The Markov semigroup (P_t) is fully specified by this infinitesimal description Q, in the sense that

- $P_t = e^{tQ}$ for all $t \geq 0$ and (equivalently)
- the forward and backward equations hold: $P_t' = P_t Q = Q P_t$.

Since $P_0 = I$, the matrix Q can be recovered from the semigroup via

$$Q = P_0' = \lim_{h \downarrow 0} \frac{P_h - I}{h}$$

In the more abstract setting of C_0 semigroups, we say that Q is the "generator" of the semigroup (P_t).

More generally, given a C_0 semigroup (U_t), we say that a linear operator A from $\mathbb{B}$ to itself is the **generator** of (U_t) if

$$Ag = \lim_{h \downarrow 0} \frac{U_h g - g}{h} \tag{6.6}$$

for all $g \in \mathbb{B}$ such that the limit exists.

The set of points where the limit exists (the domain of the generator) is denoted by $D(A)$.

At this point we would like to write (6.6) as $A = U_0'$, or express U_t as e^{tA}, analogous to the Markov case.

There are problems, however.

One problem is that the limit in (6.6) can fail to exist for some $g \in \mathbb{B}$.

Indeed, why should the limit exist, given that C_0 semigroups are not required to be differentiable?

The other problem is that, even though the limit exists, the linear operator A might be unbounded (i.e., not an element of $\mathcal{L}(\mathbb{B})$), in which case a statement like $U_t = e^{tA}$ is problematic.

It turns out that, despite these issues, the theory of C_0 semigroups is powerful, and, with some work, the technical issues can be circumvented.[5]

Even better, for the applications we wish to consider, we can focus on UC semigroups, where these problems do not arise.

The next section gives details.

---

[5] An excellent treatment of the general theory of C_0 semigroups, can be found in [Bobrowski, 2005].

6.4.3 A Characterization of Uniformly Continuous Semigroups

We saw in *Example 6.3* that exponential curves are an example of a UC semigroup.

The next theorem tells us that there are no other examples.

Theorem (UC Semigroups are Exponential Curves)

If (U_t) is a UC semigroup on $\mathbb{B}$, then there exists an $A \in \mathcal{L}(\mathbb{B})$ such that $U_t = e^{tA}$ for all $t \geq 0$. Moreover,

- U_t is differentiable at every $t \geq 0$,

- A is the generator of (U_t) and

- $U'_t = AU_t = U_t A$ for all $t \geq 0$.

The last three claims in *Theorem 6.1* follow directly from the first claim.

The statement $U'_t = AU_t = U_t A$ is a generalization of the Kolmogorov forward and backward equations.

While slightly more complicated in the Banach setting, the proof of the first claim (existence of an exponential representation) is a direct extension of the fact that any continuous function f from $\mathbb{R}$ to itself satisfying

- $f(s)f(t) = f(s + t)$ for all $s, t \geq 0$ and

- $f(0) = 1$

also satisfies $f(t) = e^{ta}$ for some $a \in \mathbb{R}$.

We proved something quite similar in *Theorem 1.1*, on the memoryless property of the exponential function.

For more discussion of the scalar case, see, for example, [Sahoo and Kannappan, 2011].

For a full proof of the first claim in *Theorem 6.1*, in the setting of a Banach algebra, see, for example, Chapter 7 of [Bobrowski, 2005].

6.5 Exercises

Exercise 6.1

Prove that (6.5) holds for all $A \in \mathcal{L}(\mathbb{B})$.

Exercise 6.2

In many texts, a C_0 semigroup is defined as an evolution semigroup (U_t) such that

$$U_t g \to g \text{ as } t \to 0 \text{ for any } g \in \mathbb{B} \tag{6.7}$$

Our aim is to show that (6.7) implies continuity at every point t, as in the definition we used above.

The Banach–Steinhaus Theorem can be used to show that, for an evolution semigroup (U_t) satisfying (6.7), there exist finite constants ω and M such that

$$\|U_t\| \leq e^{t\omega} M \quad \text{for all } t \geq 0 \tag{6.8}$$

Using this and (6.7), show that, for any $g \in \mathbb{B}$, the map $t \mapsto U_t g$ is continuous at all t.

Exercise 6.3

Following on from the previous exercise, a UC semigroup is often defined as an evolution semigroup (U_t) such that

$$\|U_t - I\| \to 0 \text{ as } t \to 0 \tag{6.9}$$

Show that (6.9) implies norm continuity at every point t, as in the definition we used above.

In particular, show that, for any $t_n \to t$, we have $\|U_{t_n} - U_t\| \to 0$ as $n \to \infty$.

6.6 Solutions

Solution to Exercise 8.1

To show the first equality, fix $t \in \mathbb{R}_+$, take $h > 0$ and observe that

$$e^{(t+h)A} - e^{tA} - e^{tA}A = e^{tA}(e^{hA} - I - A)$$

Since the norm on $\mathcal{L}(\mathbb{B})$ is submultiplicative, it suffices to show that $\|e^{hA} - I - A\| = o(h)$ as $h \to 0$.

Using the definition of the exponential, this is easily verified, completing the proof of the first equality in (6.5).

The proof of the second equality is similar.

Solution to Exercise 8.2

Let (U_t) be an evolution semigroup satisfying (6.7) and let ω and M be as in (6.8).

Pick any $g \in \mathbb{B}$, $t > 0$ and $h_n \downarrow 0$ as $n \to \infty$.

On one hand, $U_{t+h_n}g = U_{h_n}U_t g \to U_t g$ by (6.7).

On the other hand, from (6.8) and the definition of the operator norm,

$$\|U_{t-h_n}g - U_t g\| = \|U_{t-h_n}(g - U_{h_n}g)\| \le e^{(t-h_n)\omega}M\|g - U_{h_n}g\| \to 0$$

as $n \to \infty$. This completes the proof.

Solution to Exercise 8.3

The solution is similar to that of the previous exercise.

Let (U_t) be an evolution semigroup satisfying (6.9), fix $t > 0$ and take (h_n) to be a scalar sequence satisfying $h_n \downarrow 0$ as $n \to \infty$.

On one hand, $U_{t+h_n} = U_{h_n}U_t \to U_t$ by (6.9).

On the other hand, from the submultiplicative property of the operator norm and (6.8),

$$\|U_{t-h_n} - U_t\| = \|U_{t-h_n}(I - U_{h_n})\| \le e^{(t-h_n)\omega}M\|I - U_{h_n}\|$$

This converges to 0 as $n \to \infty$, completing our proof.

UC MARKOV SEMIGROUPS

7.1 Overview

In our previous lecture we covered some of the general theory of operator semigroups.

Next we translate these results into the setting of Markov semigroups.

The Markov semigroups are defined on a countable set S.

The main aim is to give an exact one-to-one correspondence between

- UC Markov semigroups

- "conservative" intensity matrices and

- jump chains with state dependent jump intensities

Conservativeness is defined below and relates to "nonexplosiveness" of the associated Markov chain.

We will also give a brief discussion of intensity matrices that do not have this property, along with the processes they generate.

7.2 Notation and Terminology

Let S be an arbitrary countable set.

Let ℓ_1 be the Banach space of **summable functions** on S; that is, all $g : S \to \mathbb{R}$ with

$$\|g\| := \sum_x |g(x)| < \infty$$

Note that $\mathcal{D}$, the set of all distributions on S, is contained in ℓ_1.

Each Markov matrix P on S can and will be identifed with a linear operator $f \mapsto fP$ on ℓ_1 via

$$(fP)(y) = \sum_x f(x)P(x,y) \qquad (f \in \ell_1, \ y \in S) \tag{7.1}$$

To be consistent with earlier notation, we are writing the argument of P to the left and applying P to it as if premultiplying P by a row vector.

In the exercises you are asked to verify that (7.1) defines a bounded linear operator on ℓ_1 such that

$$\|P\| = 1 \text{ and } \phi P \in \mathcal{D} \text{ whenever } \phi \in \mathcal{D} \tag{7.2}$$

Note that composition of P with itself is equivalent to powers of the matrix under matrix multiplication.

For an intensity matrix Q on S we can try to introduce the associated operator analogously, via

$$(fQ)(y) = \sum_x f(x)Q(x,y) \qquad (f \in \ell_1,\ y \in S) \tag{7.3}$$

However, the sum in (7.3) is not always well defined.[1]

We say that an intensity matrix Q is **conservative** if the sum in (7.3) is well defined at all y and, in addition, the mapping $f \mapsto fQ$ in (7.3) is a bounded linear operator on ℓ_1.

Below we show how this property can be checked in applications.

7.3 UC Markov Semigroups and their Generators

Let Q be a conservative intensity matrix on S.

Since Q is in $\mathcal{L}(\ell_1)$, the operator exponential e^{tQ} is well defined as an element of $\mathcal{L}(\ell_1)$ for all $t \geq 0$.

Moreover, by *Example 6.3*, the family (P_t) in $\mathcal{L}(\ell_1)$ defined by $P_t = e^{tQ}$ defines a UC Markov semigroup on ℓ_1.

(Here, a Markov semigroup (P_t) is both a collection of Markov matrices and a collection of operators, as in (7.1).)

The next theorem says that this is the only way UC Markov semigroups can arise.

Theorem 7.1

If (P_t) is a UC Markov semigroup on ℓ_1, then there exists a conservative intensity matrix Q such that $P_t = e^{tQ}$ for all $t \geq 0$.

Proof. Let (P_t) be a UC Markov semigroup on ℓ_1.

Since (P_t) is a UC semigroup on ℓ_1, it follows from *Theorem 6.1* that there exists a $Q \in \mathcal{L}(\ell_1)$ such that $P_t = e^{tQ}$ for all $t \geq 0$.

We only need to show that Q is a conservative intensity matrix.

Because (P_t) is a Markov semigroup, P_t is a Markov matrix for all t, and, since $P_t = e^{tQ}$ for all t, it follows that Q is an intensity matrix.

We proved this for the case $|S| < \infty$ in *Theorem 5.1* and one can verify that the same arguments go through when $|S| = \infty$.

As $Q \in \mathcal{L}(\ell_1)$, we know that Q is a bounded operator, so Q is a conservative intensity matrix.

From *Theorem 7.1* we can easily deduce that

- P_t is differentiable at every $t \geq 0$,
- Q is the generator of (P_t) and
- $P_t' = QP_t = P_tQ$ for all $t \geq 0$.
- $P_0' = Q$

---

[1] Previously, we introduced the notion of an intensity matrix when S is finite and the definition is essentially unchanged in the current setting. In particular, $Q : S \times S \to \mathbb{R}$ is called an **intensity matrix** if Q has zero row sums and $Q(x,y) \geq 0$ whenever $x \neq y$.

In fact these results are just a special case of the claims in *Theorem 6.1*.

The second last of these results is the Kolmogorov forward and backward equations.

The last of these results shows that we can obtain the intensity matrix Q by differentiating P_t at $t = 0$.

Example 7.1

Let us consider again the Poisson process (N_t) with rate $\lambda > 0$ in light of the discussion above.

The corresponding semigroup (P_t) is UC and hence there exists a conservative intensity matrix Q with $P_t = e^{tQ}$ for all $t \geq 0$.

This fact can be established by proving UC property and then appealing to *Theorem 7.1*.

Another alternative, easier in this case, is to supply the intensity matrix Q directly and then verify that $P_t = e^{tQ}$ holds.

The semigroup for a Poisson process with rate λ was given in (3.9) and is repeated here:

$$P_t(j, k) = \begin{cases} e^{-\lambda t} \frac{(\lambda t)^{k-j}}{(k-j)!} & \text{if } j \leq k \\ 0 & \text{otherwise} \end{cases} \tag{7.4}$$

For the intensity matrix we take

$$Q := \begin{pmatrix} -\lambda & \lambda & 0 & 0 & 0 & \cdots \\ 0 & -\lambda & \lambda & 0 & 0 & \cdots \\ 0 & 0 & -\lambda & \lambda & 0 & \cdots \\ 0 & 0 & 0 & -\lambda & \lambda & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix} \tag{7.5}$$

The form of Q is intuitive: probability flows out of state i and into state $i + 1$ at the rate λ.

It is immediate that Q is an intensity matrix, as claimed.

The exercises ask you to confirm that Q is in $\mathcal{L}(\ell_1)$.

To prove that $P_t = e^{tQ}$ for any $t \geq 0$, we first decompose Q as $Q = \lambda(K - I)$, where K is defined by

$$K(i, j) = \mathbb{1}\{j = i + 1\}$$

For given $t \geq 0$, we then have

$$e^{tQ} = e^{\lambda t(K-I)} = e^{-\lambda t} e^{\lambda t K} = e^{-\lambda t} \sum_{m \geq 0} \frac{(\lambda t K)^m}{m!}$$

The exercises ask you to verify that, for the powers of K, we have $K^m(i, j) = \mathbb{1}\{j = i + m\}$.

Inserting this expression for K^m leads to

$$e^{tQ}(i, j) = e^{-\lambda t} \sum_{m \geq 0} \frac{(\lambda t)^m}{m!} \mathbb{1}\{j = i + m\} = e^{-\lambda t} \sum_{m \geq 0} \frac{(\lambda t)^m}{m!} \mathbb{1}\{m = j - i\}$$

This is identical to (7.4).

It now follows that $t \mapsto P_t \in \mathcal{L}(\ell_1)$ is differentiable at every $t \geq 0$ and Q is the generator of (P_t), with $P_0' = Q$.

7.3.1 A Necessary and Sufficient Condition

Our definition of a conservative intensity matrix works for the theory above but can be hard to check in appliations and lacks probabilistic intuition.

Fortunately, we have the following simple characterization.

Lemma 7.1

An intensity matrix Q on S is conservative if and only if $\sup_x |Q(x,x)|$ is finite.

The proof is a solved exercise.

Example 7.2

Recall the jump chain setting where, repeating (4.4), we defined Q via

$$Q(x,y) := \lambda(x)(K(x,y) - I(x,y)) \tag{7.6}$$

The function $\lambda \colon S \to \mathbb{R}_+$ gives the jump rate at each state, while K is the Markov matrix for the embedded discrete time jump chain.

Previously we discussed this in the case where S is finite but there is no need to restrict attention to that case.

For general countable S, the matrix Q defined in (7.6) is still an intensity matrix.

If we continue to assume that $K(x,x) = 0$ for all x, then $Q(x,x) = -\lambda(x)$.

Hence, Q is conservative if and only if $\sup_x \lambda(x)$ is finite.

In other words, Q is conservative if the set of jump rates is bounded.

This example shows that requiring Q to be conservative is a relatively mild restriction.

7.3.2 The Finite State Case

It is immediate from *Lemma 7.1* that every intensity matrix is conservative when the state space S is finite.

Hence, in this setting, every intensity matrix Q on S defines a UC Markov semigroup (P_t) via $P_t = e^{tQ}$.

Conversely, if S is finite, then any Markov semigroup (P_t) is a UC Markov semigroup.

To see this, recall that, as a Markov semigroup, (P_t) satisfies $\lim_{t\to 0} P_t(x,y) = I(x,y)$ for all x,y in S.

In any finite dimensional space, pointwise convergence implies norm convergence, so $P_t \to I$ in operator norm as $t \to 0$ from above.

As we saw previously, this is enough to ensure that $t \mapsto P_t$ is norm continuous everywhere on $\mathbb{R}_+$.

Hence (P_t) is a UC Markov semigroup, as claimed.

Combining these results with *Theorem 7.1*, we conclude that, when S is finite, there is a one-to-one correspondence between Markov semigroups and intensity matrices.

7.4 From Intensity Matrix to Jump Chain

We now understand that there is a one-to-one pairing between conservative intensity matrices and UC Markov semigroups.

These ideas are important from an analytical perspective.

Now we provide another point of view, more connected to probability.

This point of view is important for both theory and computation.

7.4.1 Jump Chain Pairs

Let us agree to call (λ, K) a **jump chain pair** if λ is a map from S to $\mathbb{R}_+$ and K is a Markov matrix on S.

It is easy to verify that the matrix Q on S defined by

$$Q(x, y) := \lambda(x)(K(x, y) - I(x, y)) \tag{7.7}$$

is an intensity matrix.

(We saw in *an earlier lecture* that Q is the intensity matrix for the jump chain (X_t) built via *Algorithm 4.1* from jump chain pair (λ, K).)

As we now show, every intensity matrix admits the decomposition in (7.7) for some jump chain pair.

7.4.2 Jump Chain Decomposition

Given an intensity matrix Q, set

$$\lambda(x) := -Q(x, x) \qquad (x \in S) \tag{7.8}$$

Next we build K, first along the principle diagonal via

$$K(x, x) = \begin{cases} 0 & \text{if } \lambda(x) > 0 \\ 1 & \text{otherwise} \end{cases} \tag{7.9}$$

Thus, if the rate of leaving x is positive, we set $K(x, x) = 0$, so that the embedded jump chain moves away from x with probability one when the next jump occurs.

Otherwise, when $Q(x, x) = 0$, we stay at x forever, so x is an **absorbing state**.

Off the principle diagonal, where $x \neq y$, we set

$$K(x, y) = \begin{cases} \frac{Q(x,y)}{\lambda(x)} & \text{if } \lambda(x) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{7.10}$$

The exercises below ask you to confirm that, for λ and K just defined,

1. (λ, K) is a jump chain pair and
2. the intensity matrix Q satisfies (7.7).

We call (λ, K) the **jump chain decomposition** of Q.

We summarize in a lemma.

Lemma 7.2

A matrix Q on S is an intensity matrix if and only if there exists a jump chain pair (λ, K) such that (7.7) holds.

7.4.3 The Conservative Case

We know from *Example 7.2* that an intensity matrix Q is conservative if and only if λ is bounded.

Moreover, we saw in *Theorem 7.1* that the pairing between conservative intensity matrices and UC Markov semigroups is one-to-one.

This leads to the following result.

Theorem 7.2

On S, there exists a one-to-one correspondence between the following sets of objects:

1. The set of all jump chain pairs (λ, K) such that λ is bounded.

2. The set of all conservative intensity matrices.

3. The set of all UC Markov semigroups.

7.4.4 Simulation

In view of the preceding discussion, we have a simple way to simulate a Markov chain given any conservative intensity matrix Q.

The steps are

1. Decompose Q into a jump chain pair (λ, K).

2. Simulate via *Algorithm 4.1*.

Recalling our discussion of the Kolmogorov backward equation, we know that this produces a Markov chain with Markov semigroup (P_t) where $P_t = e^{tQ}$ for Q satisfying (7.7).

(Although our argument assumed finite S, the proof goes through when S is countably infinite and Q is conservative with very minor changes.)

In particular, (X_t) is a continuous time Markov chain with intensity matrix Q.

7.5 Beyond Bounded Intensity Matrices

If we do run into an application where an intensity matrix Q is not conservative, what might we expect?

In this scenario, we can at least hope that Q is the generator of a C_0 semigroup.

Since Q is an intensity matrix, we can be sure that this semigroup will be a Markov semigroup.

To know when Q will be the generator of a C_0 semigroup, we need to look to the Hille-Yoshida Theorem and sufficient conditions derived from it.

While we omit a detailed treatment, it is worth noting that the issue is linked to explosions.

To see the connection, recall that some initial value problems do not lead to a valid solution defined for all $t \in \mathbb{R}_+$.

An example is the scalar problem $x_t' = 1 + x_t^2$, which has solution $x_t = \tan(t - c)$ for some constant c.

But this solution equals $+\infty$ for $t \geq c + \pi/2$.

The problem is that the time path explodes to infinity in finite time.

The same issue can occur for Markov processes, if jump rates grow sufficiently quickly.

For more discussion, see, for example, Section 2.7 of [Norris, 1998].

7.6 Exercises

Exercise 7.1

Let P be a Markov matrix on S and identify it with the linear operator in (7.1). Verify the claims in (7.2).

Exercise 7.2

Prove the claim in *Lemma 7.1*.

Exercise 7.3

Confirm that Q defined in (7.5) induces a bounded linear operator on ℓ_1 via (7.3).

Exercise 7.4

Let K be defined on $\mathbb{Z}_+ \times \mathbb{Z}_+$ by $K(i,j) = \mathbb{1}\{j = i+1\}$.

Show that, with K^m representing the m-th matrix product of K with itself, we have $K^m(i,j) = \mathbb{1}\{j = i+m\}$ for any $i,j \in \mathbb{Z}_+$.

Exercise 7.5

Let Q be any intensity matrix on S.

Prove that the jump chain decomposition of Q is in fact a jump chain pair.

Prove that, in addition, this decomposition (λ, K) satisfies (7.7).

7.7 Solutions

Solution to Exercise 7.1

To determine the norm of P, we use the definition in (6.2).

If $f \in \ell_1$ and $\|f\| \le 1$, then

$$\|fP\| \le \sum_y \sum_x |f(x)| P(x,y) = \sum_x |f(x)| \sum_y P(x,y) = \sum_x |f(x)| = \|f\|$$

Hence $\|P\| \le 1$.

To see that equality holds we can repeat this argument with $f \ge 0$, obtaining $\|fP\| = \|f\|$.

Now pick any $\phi \in \mathcal{D}$.

Clearly $\phi P \geq 0$, and

$$\sum_y (\phi P)(y) = \sum_y \sum_x \phi(x) P(x, y) = \sum_x \phi(x) \sum_y P(x, y) = 1$$

Hence $\phi P \in \mathcal{D}$ as claimed.

Solution to Exercise 7.2

Here is one solution.

Let Q be an intensity matrix on S.

Suppose first that $m := \sup_x |Q(x, x)|$ is finite.

Set $\hat{P} := I + Q/m$.

It is not hard to check that $\hat{P}$ is a Markov matrix and that $Q = m(\hat{P} - I)$.

Since $\hat{P}$ is a Markov matrix, it induces a bounded linear operator on ℓ_1 via (7.1).

As $\mathcal{L}(\ell_1)$ is a linear space, we see that Q is likewise in $\mathcal{L}(\ell_1)$.

In particular, Q is a bounded operator, and hence conservative.

Next, suppose that Q is conservative and yet $\sup_x |Q(x, x)|$ is infinite.

Choose $x \in S$ such that $|Q(x, x)| > \|Q\|$

Let $f \in \ell_1$ be defined by $f(z) = \mathbb{1}\{z = x\}$.

Since $\|f\| = 1$, we have

$$\|Q\| \geq \|fQ\| = \sum_y \left| \sum_z f(z) Q(z, y) \right| = \sum_y |Q(x, y)| \geq |Q(x, x)|$$

Contradiction.

Solution to Exercise 7.3

Linearity is obvious so we focus on boundedness.

For any $f \in \ell_1$ and this choice of Q, we have

$$\sum_y |(fQ)(y)| \leq \sum_y \sum_x |f(x) Q(x, y)| \leq \lambda \sum_y \sum_x |f(y) - f(y + 1)|$$

Applying the triangle inequality, we see that the right hand side is dominated by $2\lambda \|f\|$.

Hence $\|fQ\| \leq 2\lambda \|f\|$, which implies that $Q \in \mathcal{L}(\ell_1)$ as required.

Solution to Exercise 7.4

The statement $K^m(i, j) = \mathbb{1}\{j = i + m\}$ holds by definition when $m = 1$.

Now suppose it holds at arbitrary m.

We then have, by definition of composition (matrix multiplication),

$$K^{m+1}(i, j) = \sum_n K(i, n) K^m(n, j) = \sum_n K(i, n) \mathbb{1}\{j = n + m\} = K(i, j - m)$$

Applying the definition $K(i, j) = \mathbb{1}\{j = i + 1\}$ completes verification of the claim.

Solution to Exercise 7.5

Let Q be an intensity matrix and let (λ, K) be the jump chain decomposition of Q.

Nonnegativity of λ is immediate from the definition of an intensity matrix.

To see that K is a Markov matrix we fix $x \in S$ and suppose first that $\lambda(x) > 0$.

Then

$$\sum_y K(x, y) = \sum_{y \neq x} K(x, y) = \sum_{y \neq x} \frac{Q(x, y)}{\lambda(x)} = 1$$

If, on the other hand, $\lambda(x) = 0$, then $\sum_y K(x, y) = 1$, is immediate from the definition.

As K is nonnegative, we see that K is a Markov matrix.

Thus, (λ, K) is a valid jump chain pair.

The proof that Q and (λ, K) satisfy (7.7) is mechanical and the details are omitted.

(Try working case-by-case, with $\lambda(x) = 0, x = y, \lambda(x) > 0, x = y$, etc.)

STATIONARITY AND ERGODICITY

In addition to what's in Anaconda, this lecture will need the following libraries:

```
!pip install quantecon
```

8.1 Overview

In this lecture we discuss stability and equilibrium behavior for continuous time Markov chains.

To give one example of why this theory matters, consider queues, which are often modeled as continuous time Markov chains.

Queueing theory is used in applications such as

- treatment of patients arriving at a hospital

- optimal design of manufacturing processes

- requests to a file server

- air traffic

- customers waiting on a helpline

A key topic in queueing theory is average behavior over the long run.

- Will the length of the queue grow without bounds?

- If not, is there some kind of long run equilibrium?

- If so, what is the average waiting time in this equilibrium?

- What is the average length of the queue over a week, or a month?

We will use the following imports

```python
import numpy as np
import scipy as sp
import matplotlib.pyplot as plt
import quantecon as qe
from numba import njit
from scipy.linalg import expm

from matplotlib import cm
from mpl_toolkits.mplot3d import Axes3D
from mpl_toolkits.mplot3d.art3d import Poly3DCollection
```

8.2 Stationary Distributions

8.2.1 Definition

Let S be countable.

Recall that, for a discrete time Markov chain with Markov matrix P on S, a distribution ψ is called stationary for P if $\psi P = \psi$.

This means that if X_t has distribution ψ, then so does X_{t+1}.

For continuous time Markov chains, the definition is analogous.

Given a Markov semigroup (P_t) on S, a distribution $\psi^* \in \mathcal{D}$ is called **stationary** for (P_t) if
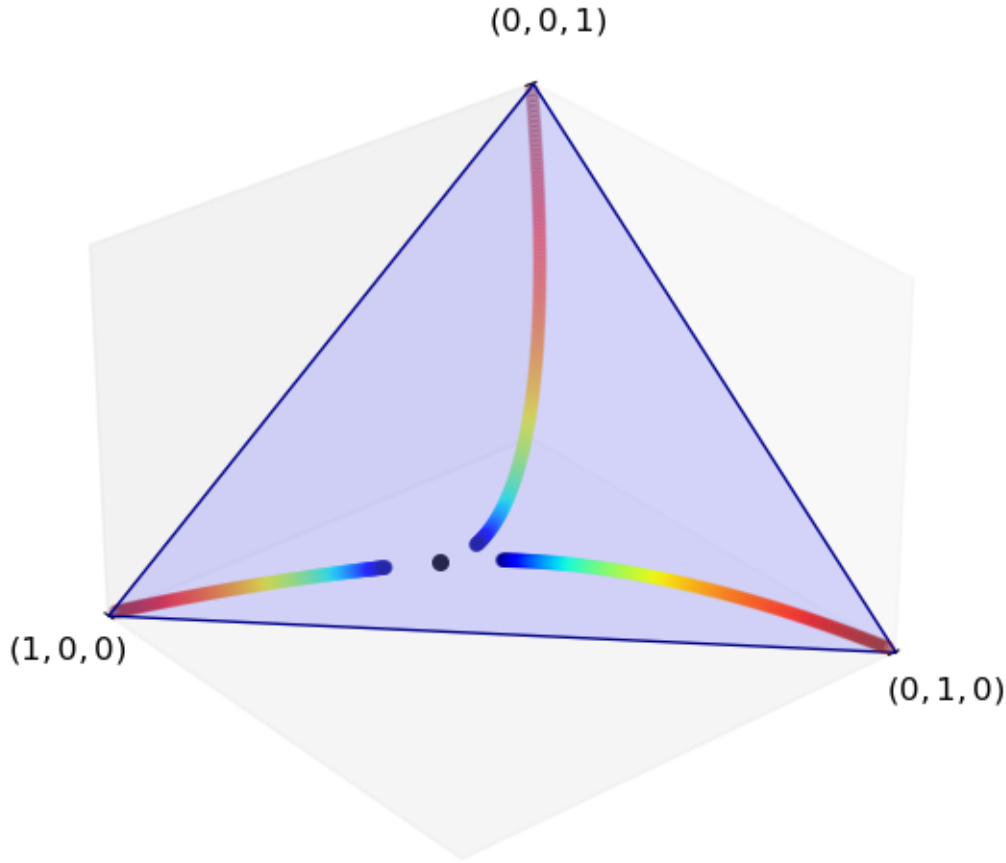
$$\psi^* P_t = \psi^* \text{ for all } t \geq 0$$

As one example, we look *again* at the chain on $S = \{0, 1, 2\}$ with intensity matrix

```
Q = ((-3,  2,  1),
     (3,  -5,  2),
     (4,   6, -10))
```

The following figure was shown before, except that now there is a black dot that the three trajectories seem to be converging to.

(Recall that, in the color scheme, trajectories cool as time evolves.)

This black dot is the stationary distribution ψ^* of the Markov semigroup (P_t) generated by Q.

It was calculated using the `stationary_distributions` attribute of QuantEcon's `MarkovChain class`, by arbitrarily setting $t = 1$ and solving $\psi P_1 = \psi$.

Below we show that, for this choice of Q, the stationary distribution ψ^* is unique in $\mathcal{D}$, due to irreducibility.

Moreover, $\psi P_t \to \psi^*$ as $t \to \infty$ for any $\psi \in \mathcal{D}$, as suggested by the figure.

8.2.2 Stationarity via the Generator

In many cases, it is easier to use the generator of the semigroup to identify stationary distributions rather than the semigroup itself.

This is analogous to the idea that a point $\bar{x}$ in $\mathbb{R}^d$ is stationary for a vector ODE $x'_t = F(x_t)$ when $F(\bar{x}) = 0$.

(Here F is the infinitesimal description, and hence analogous to the generator.)

The next result holds true under weaker conditions but the version stated here is easy to prove and sufficient for applications we consider.

Theorem 8.1

Let (P_t) be a UC Markov semigroup with generator Q. A distribution ψ on S is stationary for (P_t) if and only if $\psi Q = 0$.

Proof. Fix $\psi \in \mathcal{D}$ and suppose first that $\psi Q = 0$.

Since (P_t) is a UC Markov semigroup, we have $P_t = e^{tQ}$ for all t, and hence, for any $t \geq 0$,

$$\psi e^{tQ} = \psi + t\psi Q + t^2 \frac{\psi Q^2}{2!} + \cdots$$

From $\psi Q = 0$ we get $\psi Q^k = 0$ for all $k \in \mathbb{N}$, so the last display yields $\psi P_t = \psi$.

Hence ψ is stationary for (P_t).

Now suppose that ψ is stationary for (P_t) and set $D_t := (1/t)(P_t - I)$.

From the triangle inequality and the definition of the operator norm, for any given t,

$$\|\psi Q\| \leq \|\psi(Q - D_t)\| + \|\psi D_t\| \leq \|Q - D_t\| + \|\psi D_t\|$$

Since (P_t) is UC and Q is its generator, we have $\|D_t - Q\| \to 0$ in $\mathcal{L}(\ell_1)$ as $t \to 0^+$.

Hence $\|\psi Q\| \leq \liminf_{t \downarrow 0} \|\psi D_t\|$.

As ψ is stationary for (P_t), we have $\psi D_t = 0$ for all t.

Hence $\psi Q = 0$, as was to be shown.

8.3 Irreducibility and Uniqueness

Let (P_t) be a Markov semigroup on S and consider arbitrary states $x, y \in S$.

We say that state y is **accessible** from state x if there exists a $t \geq 0$ such that $P_t(x, y) > 0$.

We say that x and y **communicate** if x is accessible from y and y is accessible from x.

A Markov semigroup (P_t) on S is called **irreducible** if every pair x, y in S communicates.

We seek a characterization of irreducibility of (P_t) in terms of its generator.

As a first step, we will say there is a Q**-positive probability flow** from x to y if there exists a finite sequence $(z_i)_{i=0}^m$ in S starting at $x = z_0$ and ending at $y = z_m$ such that $Q(z_i, z_{i+1}) > 0$ for all i.

Theorem 8.2

Let (P_t) be a UC Markov semigroup with generator Q. For distinct states x and y, the following statements are equivalent:

1. The state y is accessible from x under (P_t).

2. There is a Q-positive probability flow from x to y.

3. $P_t(x, y) > 0$ for all $t > 0$.

Proof. Pick any two distinct states x and y.

It is obvious that statement 3 implies statement 1, so we need only prove $(1 \implies 2)$ and $(2 \implies 3)$.

Starting with $(1 \implies 2)$, recall that

$$P_t(x, y) = tQ(x, y) + \frac{t^2}{2!}Q^2(x, y) + \cdots \tag{8.1}$$

If x is accessible from y, then $P_t(x, y) > 0$ for some $t > 0$, so $Q^k(x, y) > 0$ for at least one $k \in \mathbb{N}$.

Writing out the matrix product as a sum, we now have

$$Q^k(x, y) = \sum_{z_1} \sum_{z_2} \cdots \sum_{z_{k-1}} Q(x, z_1) Q(z_1, z_2) \cdots Q(z_{k-1}, y) > 0 \tag{8.2}$$

It follows that at least one element of the sum must be strictly positive.

Therefore, a Q-positive probability flow from x to y exists.

Turning to ($2 \implies 3$), first note that, for arbitrary states u and v, if $Q(u, v) > 0$ then $P_t(u, v) > 0$ for all $t > 0$.

To see this, let (λ, K) be the jump chain pair constructed from Q via (7.8), (7.9) and (7.10).

Observe that, since $Q(u, v) > 0$, we must have $\lambda(u) > 0$.

As a consequence, applying (7.10), we have

$$K(u, v) = \frac{Q(u, v)}{\lambda(u)} > 0$$

Let (Y_k) and (J_k) be the embedded jump chain and jump sequence generated by *Algorithm 4.1*, with $Y_0 = u$.

With $E_1 \sim \text{Exp}(1)$ and $E_2 \sim \text{Exp}(1)$, we have, for any $t > 0$,

$$\begin{aligned}
P_t(u, v) &\geq \mathbb{P}\{J_1 \leq t,\ Y_1 = v,\ J_2 > t\} \\
&\geq \mathbb{P}\{E_1 \leq t\lambda(u),\ E_2 > t\lambda(v)\}\mathbb{P}\{Y_1 = v\} \\
&= \mathbb{P}\{E_1 \leq t\lambda(u)\}\mathbb{P}\{E_2 > t\lambda(v)\}K(u, v) \\
&> 0
\end{aligned}$$

Now suppose there is a Q-positive probability flow $(z_i)_{i=0}^m$ from x to y.

If we fix $t > 0$ and repeatedly apply (3.7) along with the last result, we obtain

$$P_t(x, y) \geq \prod_{i=0}^{m-1} P_{t/m}(z_i, z_{i+1}) > 0$$

Theorem 8.2 leads directly to the following strong result.

Corollary 8.1

For a UC Markov semigroup (P_t), the following statements are equivalent:

1. (P_t) is irreducible.
2. $P_t(x, y) > 0$ for all $t > 0$ and all $(x, y) \in S \times S$.

Note: To obtain stable long run behavior in discrete time Markov chains, it is common to assume that the chain is aperiodic.

This needs to be assumed on top of irreducibility if one wishes to rule out all dependence on initial conditions.

Corollary 8.1 shows that periodicity is not a concern for irreducible continuous time Markov chains.

Positive probability flow from x to y at some $t > 0$ immediately implies positive flow for all $t > 0$.

8.4 Asymptotic Stabilitiy

We call Markov semigroup (P_t) **asymptotically stable** if (P_t) has a unique stationary distribution ψ^* in $\mathcal{D}$ and

$$\|\psi P_t - \psi^*\| \to 0 \text{ as } t \to \infty \text{ for all } \psi \in \mathcal{D} \tag{8.3}$$

Our aim is to establish conditions for asymptotic stability of Markov semigroups.

8.4.1 Contractivity

Let's recall some useful facts about the discrete time case.

First, if P is any Markov matrix, we have, in the ℓ_1 norm,

$$\|\psi P\| \leq \|\psi\| \text{ for all } \psi \in \mathcal{D} \tag{8.4}$$

This is because, given $\psi \in \mathcal{D}$,

$$\|\psi P\| = \sum_y \left| \sum_x \psi(x) P(x, y) \right| \leq \sum_y \sum_x |\psi(x)| P(x, y) = \|\psi\|$$

By linearity, for $\psi, \phi \in \mathcal{D}$, we then have

$$\|\psi P - \phi P\| \leq \|\psi - \phi\|$$

Hence every Markov operator is contracting on $\mathcal{D}$.

Moreover, if P is everywhere positive, then this inequality is strict:

Lemma 8.1 (Strict Contractivity)

If P is a Markov matrix and $P(x, y) > 0$ for all x, y, then

$$\|\psi P - \phi P\| < \|\psi - \phi\| \text{ for all } \psi, \phi \in \mathcal{D} \text{ with } \psi \neq \phi$$

The proof follows from the strict triangle inequality, as opposed to the weak triangle inequality we used to obtain (8.4).

See, for example, Proposition 3.1.2 of [Lasota and Mackey, 1994] or Lemma 8.2.3 of [Stachurski, 2009].

8.4.2 Uniqueness

Irreducibility of a given Markov chain implies that there are no disjoint absorbing sets.

This in turn leads to uniqueness of stationary distributions:

Theorem 8.3

Let (P_t) be a UC Markov semigroup on S. If (P_t) is irreducible, then (P_t) has at most one stationary distribution.

 Proof. Suppose to the contrary that ψ and ϕ are both stationary for (P_t).

Since (P_t) is irreducible, we know that $P_1(x, y) > 0$ for all $x, y \in S$.

If $\psi \neq \phi$, then, due to positivity of P_1, the strict inequality in *Lemma 8.1* holds.

At the same time, by stationarity, $\|\psi P - \phi P\| = \|\psi - \phi\|$. Contradiction.

Example 8.1

An M/M/1 queue with parameters μ, λ is a continuous time Markov chain (X_t) on $S = \mathbb{Z}_+$ with with intensity matrix

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \cdots \\ \mu & -(\mu + \lambda) & \lambda & 0 & \cdots \\ 0 & \mu & -(\mu + \lambda) & \lambda & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{8.5}$$

The chain (X_t) records the length of the queue at each moment in time.

The intensity matrix captures the idea that customers flow into the queue at rate λ and are served (and hence leave the queue) at rate μ.

If λ and μ are both positive, then there is a Q-positive probability flow between any two states, in both directions, so the corresponding semigroup (P_t) is irreducible.

Theorem 8.3 now tells us that (P_t) has at most one stationary distribution.

8.4.3 Stability from the Skeleton

Recall the definition of asymptotic stability given in (8.3).

Analogously, we call an individual Markov operator P asymptotically stable if P has a unique stationary distribution ψ^* in $\mathcal{D}$ and $\psi P^n \to \psi^*$ as $n \to \infty$ for all $\psi \in \mathcal{D}$.

The next result gives a connection between discrete and continuous stability.

The critical ingredient linking these two concepts is the contractivity in (8.4).

Lemma 8.2

Let (P_t) be a Markov semigroup. If there exists an $s > 0$ such that the Markov matrix P_s is asymptotically stable, then (P_t) is asymptotically stable with the same stationary distribution.

Proof. Let (P_t) and s be as in the statement of *Lemma 8.2*.

Let ψ^* be the stationary distribution of P_s. Fix $\psi \in \mathcal{D}$ and $\epsilon > 0$.

By stability of P_s, we can take an $n \in \mathbb{N}$ such that $\|\psi P_s^n - \psi^*\| < \epsilon$.

Pick any $t > sn$. Set $h := t - sn$.

By the contractivity in (8.4) and $P_{sn} = P_s^n$, we have

$$\|\psi P_t - \psi^*\| = \|\psi P_{sn} P_h - \psi^* P_h\| \leq \|\psi P_{sn} - \psi^*\| < \epsilon$$

Hence asymptotic stability holds for (P_t).

8.4.4 Stability via Drift

In this section we address drift conditions, which are a powerful method for obtaining asymptotic stability when the state space can be infinite.

The idea is to show that the state tends to drift back to a finite set over time.

Such drift, when combined with the contractivity in *Lemma 8.1*, is enough to give global stability.

The next theorem gives a useful version of this class of results.

Theorem 8.4

Let (P_t) be a UC Markov semigroup with intensity matrix Q. If (P_t) is irreducible and there exists a function $v \colon S \to \mathbb{R}_+$, a finite set $F \subset S$ and positive constants ϵ and M such that

$$\sum_y Q(x, y)v(y) \leq \begin{cases} M & \text{if } x \in F \text{ and} \\ -\epsilon & \text{otherwise} \end{cases}$$

then (P_t) is asymptotically stable.

The proof of *Theorem 8.4* can be found in [Pichór *et al.*, 2012].

Example 8.2

Consider again the M/M/1 queue on $\mathbb{Z}_+$ with intensity matrix (8.5).

Suppose that $\lambda < \mu$.

It is intuitive that, in this case, the queue length will not tend to infinity (since the service rate is higher than the arrival rate).

This intuition can be confirmed via *Theorem 8.4*, after setting $v(j) = j$.

Indeed, we have, for any $i \geq 1$,

$$\sum_{j \geq 0} Q(i, j)v(j) = (i - 1)\mu - i(\mu + \lambda) + (i + 1)\lambda = \lambda - \mu$$

Setting $F = \{0\}$ and $M = \lambda - \mu = -\epsilon$, we see that the conditions of *Theorem 8.4* hold.

Hence the associated semigroup (P_t) is asymptotically stable.

Corollary 8.2

If (P_t) is an irreducible UC Markov semigroup and S is finite, then (P_t) is asymptotically stable.

A solved exercise below asks you to confirm this.

8.5 Exercises

Exercise 8.1

Let (P_t) be a Markov semigroup. True or false: for this semigroup, every state x is accessible from itself.

Exercise 8.2

Let (λ_k) be a bounded non-increasing sequence in $(0, \infty)$.

A **pure birth process** starting at zero is a continuous time Markov process (X_t) on state space $\mathbb{Z}_+$ with intensity matrix

$$Q = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \cdots \\ 0 & -\lambda_1 & \lambda_1 & 0 & \cdots \\ 0 & 0 & -\lambda_2 & \lambda_2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Show that (P_t), the corresponding Markov semigroup, has no stationary distribution.

Exercise 8.3

Confirm that *Theorem 8.4* implies *Corollary 8.2*.

8.6 Solutions

Solution to Exercise 8.1

The statement is true. With $t = 0$ we have $P_t(x, x) = I(x, x) = 1 > 0$.

Solution to Exercise 8.2

Suppose to the contrary that $\phi \in \mathcal{D}$ and $\phi Q = 0$.

Then, for any $j \geq 1$,

$$(\phi Q)(j) = \sum_{i \geq 0} \phi(i) Q(i, j) = -\lambda_j \phi(j) + \lambda_{j-1} \phi(j-1) = 0$$

Since (λ_k) is non-increasing, it follows that

$$\frac{\phi(j)}{\phi(j-1)} = \frac{\lambda_{j-1}}{\lambda_j} \geq 1$$

Therefore, for any $j \geq 1$, it must be:

$$\phi(j) \geq \phi(j-1)$$

It follows that ϕ is non-decreasing on $\mathbb{Z}_+$.

But $\mathcal{D}$ contains no non-decreasing functions when the state space is infinite. (Why?)

Contradiction.

Solution to Exercise 8.3

Let (P_t) be an irreducible UC Markov semigroup and let S be finite.

Pick any positive constants M, ϵ and set $v = M$ and $F = S$.

We then have

$$\sum_y Q(x, y)v(y) = M \sum_y Q(x, y) = 0$$

Hence the drift condition in *Theorem 8.4* holds and (P_t) is asymptotically stable.

BIBLIOGRAPHY

9.1 References

BIBLIOGRAPHY

[App19]   David Applebaum. *Semigroups of Linear Operators*. Volume 93. Cambridge University Press, 2019.

[Bob05]   Adam Bobrowski. *Functional analysis for probability and stochastic processes: an introduction*. Cambridge University Press, 2005.

[How17]   Douglas C Howard. *Elements of Stochastic Processes: A Computational Approach*. FE Press, 2017.

[LM94]   Andrzej Lasota and Michael C Mackey. *Chaos, fractals, and noise: stochastic aspects of dynamics*. Volume 97. Springer Science & Business Media, 1994.

[LG16]   Jean-François Le Gall. *Brownian motion, martingales, and stochastic calculus*. Volume 274. Springer, 2016.

[Lig10]   Thomas Milton Liggett. *Continuous time Markov processes: an introduction*. Volume 113. American Mathematical Soc., 2010.

[Nor98]   James R Norris. *Markov chains*. Number 2. Cambridge University Press, 1998.

[Par08]   Etienne Pardoux. *Markov processes and applications: algorithms, networks, genome and finance*. Volume 796. John Wiley & Sons, 2008.

[PichorRTKaminska12]   Katarzyna Pichór, Ryszard Rudnicki, and Marta Tyran-Kamińska. Stochastic semigroups and their applications to biological models. *Demonstratio Mathematica*, 45(2):463–494, 2012.

[SK11]   Prasanna K Sahoo and Palaniappan Kannappan. *Introduction to functional equations*. CRC Press, 2011.

[Sta09]   John Stachurski. *Economic dynamics: theory and computation*. MIT Press, 2009.

[Str13]   Daniel W Stroock. *An introduction to Markov processes*. Volume 230. Springer Science & Business Media, 2013.

[Wal12]   John B Walsh. *Knowing the odds: an introduction to probability*. Volume 139. American Mathematical Soc., 2012.

theorem-5

uc-mc-semigroups-prf-1

ucsgec

uniirr

usmg