Advanced Macro
Fall 2011

Miklós Koren
korenm@ceu.hu
koren.hu/teaching

Lecture 11: Continuous time dynamics

Introduction

Goal of the course

- In the next 5 classes, we will study continuous-time dynamic programming.

Goal of the course

- In the next 5 classes, we will study continuous-time dynamic programming.
- By the end of the course, I would like you to feel passionate about:
  - dynamic programming
  - continuous-time modelling

Learning outcomes

You will have applicable knowledge of

- discrete-state Markov processes in continuous time
- dynamic programming in continuous time
  - without uncertainty
  - with discrete-state uncertainty
- using dynamic programming in general equilibrium
- aggregation of heterogeneous agent models

Applications

We will cover three applications:

1. The expanding variety model of growth (Grossman and Helpman)
2. The rising product quality model of growth (Aghion and Howitt)
3. A firm-level model of innovation (Klette and Kortum)

Outline

Outline

- Today we review Markov processes.
- We show how they work in continuous time.
- We consider two cases:
  - no uncertainty
  - discrete states

What we don't do

- Hamiltonians and calculus of variations

What we don't do

- Hamiltonians and calculus of variations
  - Hamiltonians are an engineer's way of thinking.
  - The recursive formulation is much more intuitive from the point-of-view of a decision maker in an ever-changing environment.

What we don't do

- Hamiltonians and calculus of variations
  - Hamiltonians are an engineer's way of thinking.
  - The recursive formulation is much more intuitive from the point-of-view of a decision maker in an ever-changing environment.
- Wiener processes and Brownian motions

What we don't do

- Hamiltonians and calculus of variations
  - Hamiltonians are an engineer's way of thinking.
  - The recursive formulation is much more intuitive from the point-of-view of a decision maker in an ever-changing environment.
- Wiener processes and Brownian motions
  - They are very useful in finance.
  - but they require a special set of tools.

Why continuous time?

Why continuous time?

- Time *is* continuous, only measurement is discrete.
    - Q1 GDP measures all the value added in the economy between January 1, 2009, 12am and March 31, 2009, 11.59.59pm.
    - Prices are measured monthly, unemployment is reported weekly.
    - Full-population census is usually done every 10 years.

Why continuous time?

- Time *is* continuous, only measurement is discrete.
  - Q1 GDP measures all the value added in the economy between January 1, 2009, 12am and March 31, 2009, 11.59.59pm.
  - Prices are measured monthly, unemployment is reported weekly.
  - Full-population census is usually done every 10 years.
- It is often useful to think about the "true" model first and then ask how it is measured.

Why continuous time?

- Time *is* continuous, only measurement is discrete.
  - Q1 GDP measures all the value added in the economy between January 1, 2009, 12am and March 31, 2009, 11.59.59pm.
  - Prices are measured monthly, unemployment is reported weekly.
  - Full-population census is usually done every 10 years.
- It is often useful to think about the "true" model first and then ask how it is measured.
- Often, continuous-time math is simpler.

Continuous time

- In continuous time, $t \in \mathbb{R}$.
- There are no special points or intervals, all t are similar.
  - We can define arbitrary intervals as we wish.
- Continuous time forces you to think about flows and stocks carefully.

Example 1: "Bill Gates could buy Costa Rica"

- *Forbes* reports that Bill Gates' net worth, \$50 billion, is higher than the GDP of Costa Rica, hence "Bill Gates could buy Costa Rica".
- We, economists, know this is totally stupid: net worth is a *stock*, GDP is a *flow*.
- But just in case:
  - In continuous time, the two actually have different units: GDP is measured in \$/year (or second), net worth is measured in \$.
  - They cannot be added, subtracted or compared.
  - Even the math does not let you commit this silly mistake.

Example 2: Cash-in-advance models

- In cash-in-advance models, all your period consumption has to be financed by your cash in hand:

$$C_t \leq M_t.$$

Example 2: Cash-in-advance models

- In cash-in-advance models, all your period consumption has to be financed by your cash in hand:

$$C_t \leq M_t.$$

- But C_t is a flow, M_t is a stock!
- This comparison does not make sense until you know how often you can replenish your stock of M.

$$C(t, t+\Delta) \equiv \int_t^{t+\Delta} c(s)ds \leq M(t)$$

- The choice of time period, Δ, is crucial.

Example 2 (continued)

- In practice, there are very few actual flows (maybe your electricity consumption).
- Income, production, consumption etc mostly happen in chunks (you rarely buy a new computer).
- We will also learn tools to deal with these rare occurrences.

When to use continuous-time modelling?

- Use continuous time if
  - you need simple and clean formulas
  - you want to think about your problem at different intervals

When to use continuous-time modelling?

- Use continuous time if
  - you need simple and clean formulas
  - you want to think about your problem at different intervals
- Use discrete time instead if
  - you want to simulate your model in a computer (for a computer, nothing is continuous)
  - you want to estimate your model on data measured at discrete intervals (years, quarters, months)
  - your model assigns a special role to certain points or intervals in time (e.g. trading day).

Markov processes

Markov processes

- A Markov process is a stochastic process for which conditional on the present state of the system, its past and future are independent.
- Time homogeneous Markov processes:

$$\Pr\big[X(t+h) = y \mid X(t) = x\big] = \Pr\big[X(h) = y \mid X(0) = x\big]$$

- Many processes have a Markovian representation.

Example: An AR(1) process

- Suppose GDP follows an AR(1) process:

$$y_t = \rho y_{t-1} + u_t$$

- Knowing y_t helps you predict y_{t+1}, y_{t+2}, etc.

Example: An AR(1) process

- Suppose GDP follows an AR(1) process:

$$y_t = \rho y_{t-1} + u_t$$

- Knowing y_t helps you predict y_{t+1}, y_{t+2}, etc.
- But the key is that *nothing else does*.

Markovian representations

- What if the future depends on the past, not only the present?
  - Say, unemployment next week depends on last week's number, but also on seasonality.
  - We can always increase the state space to include last year's number.

Markovian representations

- What if the future depends on the past, not only the present?
  - Say, unemployment next week depends on last week's number, but also on seasonality.
  - We can always increase the state space to include last year's number.
- What if there are "news" about the future that are informative?
  - Say, technological developments are reported in Science and this affects future GDP growth.
  - Again, we can include a state variable (or vector) to account for these news.
  - This works as long the news themselves have Markovian dynamics.

Markovian representations

- What if the future depends on the past, not only the present?
  - Say, unemployment next week depends on last week's number, but also on seasonality.
  - We can always increase the state space to include last year's number.
- What if there are "news" about the future that are informative?
  - Say, technological developments are reported in Science and this affects future GDP growth.
  - Again, we can include a state variable (or vector) to account for these news.
  - This works as long the news themselves have Markovian dynamics.
- One can use Markovian tools for even inherently forward-looking phenomena.
  - Say, the continuation value in a dynamic contract can be a state variable. (Sargent calls this "dynamic programming squared".)

No uncertainty

Discrete time review

First-order difference equations

- Let x_t be a $k \times 1$ vector.
- x_t follows a first-order difference equation if $\Delta x_t \equiv x_t - x_{t-1}$ is a function of x_{t-1}:

$$\Delta x_t = F(x_{t-1}).$$

  - This is a special case of a first-order Markov process.
  - Being first order is not restrictive. Why?
- Long-run stability, speed of converge etc. can be characterized by certain properties of F.

Cobweb plot

Continuous time

Moving to continuous time

- Let time periods be Δ apart.
- How can we characterize the time series as Δ becomes smaller and smaller?
- We take the limit as $\Delta \to 0$.
  - Often, we will have to rescale changes in the variable for the limit to make sense.

Differential equations

- Suppose x follows a difference equation

$$x_{t+\Delta} - x_t = F(x_t, \Delta).$$

- Note that F may depend on Δ.
  - It is unreasonable to assume that the equation of motion is the same for a day as for a year.
- Let's look at the *rate* of change in x:

$$\frac{x_{t+\Delta} - x_t}{\Delta} = \frac{F(x_t, \Delta)}{\Delta}.$$

Infinitesimal changes

- Now let $\Delta \to 0$:

$$\lim_{\Delta \to 0} \frac{x_{t+\Delta} - x_t}{\Delta} \equiv \frac{dx(t)}{dt} \equiv \dot{x}(t)$$

$$\dot{x}(t) = \lim_{\Delta \to 0} \frac{F(x_t, \Delta)}{\Delta} \equiv f(x_t).$$

- Note that for $f(x_t)$ to exist, $F(x_t, \Delta)$ has to be $O(\Delta)$.
  - Detour: O, o notation.

Infinitesimal changes

- Now let $\Delta \to 0$:

$$\lim_{\Delta \to 0} \frac{x_{t+\Delta} - x_t}{\Delta} \equiv \frac{dx(t)}{dt} \equiv \dot{x}(t)$$

$$\dot{x}(t) = \lim_{\Delta \to 0} \frac{F(x_t, \Delta)}{\Delta} \equiv f(x_t).$$

- Note that for $f(x_t)$ to exist, $F(x_t, \Delta)$ has to be $O(\Delta)$.
  - Detour: O, o notation.
- Dynamics is described by the above ordinary differential equation.
- Often, we cannot solve for $x(t)$ in closed form.
- We can still characterize the steady state, its stability, speed of convergence by the properties of f.

Steady state

- The steady state of this system is x^* such that

$$f(x^*) = 0.$$

Stability

- The local stability of the steady state depends on the derivative (*gradient*) of f.
- (See Katrin's math course for details.)

Example 1: The Solow model

- The law of motion for capital in the Solow model:

$$\dot{k} = sf(k) - \delta k.$$

- The steady-state capital is implicitly given by

$$sf(k^*) = \delta k^*.$$

- The steady state is stable if f is concave.

Phase diagram

Example 2: The Ramsey model

- The law of motion for capital and consumption in the Ramsey model (we will derive it later):

$$\dot{k} = f(k) - \delta k - c$$
$$\dot{c} = \frac{f'(k) - \delta - \rho}{\theta} c$$

Phase diagram

Uncertainty

Discrete time review

Markov chains

- Let x_t take one of N discrete values: $\{S_1, ..., S_N\}$.
- We denote the probability that $x_t = S_n$ by π_{nt}.
- The row vector of probabilities is $\pi_t = \{\pi_{1t}, ..., \pi_{Nt}\}$.
- The probability of moving from state i to state j is P_{ij}.
- These probabilities can be collected in a *transition matrix* P.

$$\pi_{t+1} = \pi_t P$$

Forecasting with Markov chains

- If the system starts from state π_0, the probability distribution of the states at time t:

$$\pi_t = \pi_0 P^t$$

- We can use this iteration to forecast any future state of the system.

Stationary distribution

- The stationary (invariant) distribution is such that

$$\pi^* = \pi^* P.$$

- This is the eigenvector corresponding to the eigenvalue 1.

Stationary distribution

- Because P is stochastic, there is always at least one eigenvalue of 1.
- π^* is unique if 1 is not a multiple eigenvalue.
- π^* is asymptotic stationary if all other eigenvalues are less than 1 in absolute value.

Example

- Take the following 2×2 transition matrix:

$$P = \begin{bmatrix} 0.9 & 0.1 \\ 0.3 & 0.7 \end{bmatrix}$$

- The steady-state distribution is

$$\pi^* = \begin{pmatrix} 0.75 \\ 0.25 \end{pmatrix}$$

Continuous time

Moving to continuous time

- Let us now take Δ to 0.
- The set of possible states is given, $S_1, ..., S_N$.
- What changes with Δ is the transition matrix: $P(\Delta)$.

Moving to continuous time

- Intuitively, we don't expect any change over a very short $\Delta \approx 0$ period of time:

$$\lim_{\Delta \to 0} P(\Delta) = I.$$

- But over any positive amount of time, the transition matrix should be nontrivial

$$P(\Delta) \neq I \text{ for all } \Delta > 0.$$

- As $\Delta \to 0$, we should put more and more weight on the diagonal elements of P (the probability of no change).
- How to describe such a Markov process?
  - Formally, we want $P(\Delta) - I$ to be $O(\Delta)$.

A rescaled transition matrix

- The probability of staying in state i:

$$P_{ii}(\Delta) = 1 - \Lambda_i \Delta.$$

- The probability of jumping to state j:

$$P_{ij}(\Delta) = \lambda_{ij} \Delta.$$

- (In fact, any function of order $O(\Delta)$ will do.)

Transitions

- So that $P(\Delta)$ is a transition matrix, we need

$$\Lambda_i = \sum_{j \neq i} \lambda_{ij}$$

- The λ_{ij}s fully characterize the dynamics of this system.
- These can be called arrival rates, hazard rates, birth rates if $j = i + 1$, death rates if $j = i - 1$.
- Note that λ_{ij} is an arrival *rate*, not a probability.
  - It can take any non-negative value, not just $[0, 1]$.

The transition "rate" matrix

- More generally, we know that $P(\Delta) - I$ is $O(\Delta)$.
- This means that
$$\lim_{\Delta \to 0} \frac{P(\Delta) - I}{\Delta} \equiv \Lambda$$
exists.
  - Because $P(\Delta)\mathbf{1} = \mathbf{1}$ (P is stochastic), $\Lambda\mathbf{1} = \mathbf{0}$.
  - The diagonal elements are negative,
  - the off-diagonals are positive.
- We call the matrix Λ the *transition rate matrix*.
  - This fully characterizes the continuous-time Markov chain.

Example 1: Employment and unemployment

- The hazard rate of losing a job is δ.
  - The lifetime of a job is exponential with mean $1/\delta$.
  - Job loss is memoryless: you are just as likely to get fired on your 2nd day as on your 366th.
- The arrival rate of a new job for an unemployed is λ.
  - The spell of unemployment is exponential with mean $1/\lambda$.
  - Unemployment is memoryless: you are just as likely to find a job after 1 day of unemployment as after 365.

Example 1: continued

- State 0: employment.
- State 1: unemployment.

$$\Lambda = \begin{bmatrix} -\delta & \delta \\ \lambda & -\lambda \end{bmatrix}$$

Example 2: incoming emails

- Let $n(t)$ be the number of emails in your inbox at time t.
- We want to characterize the dynamics of n.
- Suppose emails arrive at random (think of spam).
  - You never erase email:

$$\lambda_{i,j} = 0 \text{ if } j < i$$

  - No two emails arrive at the same time:

$$\lambda_{i,i+s} = 0 \text{ for all } s \geq 2$$

  - Each new email arrives with a constant arrival rate:

$$\lambda_{i,i+1} = \lambda$$

  - By construction, $\Lambda_i = \lambda$.
- This is called a Poisson process.

Example 2 (continued)

The transition rate matrix for the Poisson process:

$$\begin{bmatrix} -\lambda & \lambda & 0 & \cdots \\ 0 & -\lambda & \lambda & \cdots \\ 0 & 0 & -\lambda & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Forecasting with Markov chains

Example 1: Employment and unemployment

- Suppose you start out employed at $t = 0$, $\pi_0(0) = 1$.
- What is the probability that you will be employed at $t + \Delta$?

$$\pi_0(t + \Delta) = (1 - \delta\Delta)\pi_0(t) + \lambda\Delta\pi_1(t)$$

  - with prob $(1 - \delta\Delta)$ you remained in state 0
  - with prob $\lambda\Delta$ you exited from state 1

Example 1: Employment and unemployment

$$\pi_0(t+\Delta) = (1-\delta\Delta)\pi_0(t) + \lambda\Delta\pi_1(t)$$

- Subtract $\pi_0(t)$ from both sides and divide by Δ.

$$\frac{\pi_0(t+\Delta)-\pi_0(t)}{\Delta} = -\delta\pi_0(t)+\lambda\pi_1(t) = -\delta\pi_0(t)+\lambda[1-\pi_0(t)]$$

- Take $\Delta \to 0$

$$\dot{\pi}_0(t) = \lambda - (\lambda+\delta)\pi_0(t)$$

- This is a first-order ordinary differential equation, known as the Kolmogorov forward equation.

The Kolmogorov equation

$$\dot{\pi}_0(t) = \lambda - (\lambda + \delta)\pi_0(t)$$

- To get the likelihood of each state at any t, we can solve the Kolmogorov equation forward, starting from an initial value at $t = 0$.
- The solution to this ODE

$$\pi_0(t) = \frac{\lambda}{\lambda + \delta}\left[1 - Ce^{-(\lambda+\delta)t}\right],$$

where C is a constant of integration (pinned down by the boundary condition).

The steady-state distribution

- The steady-state distribution of unemployment is

$$\lim_{t \to \infty} \pi_0(t) = \frac{\lambda}{\lambda + \delta}.$$

- (We will also derive with another method.)

More generally

- More generally, the Kolmogorov equation is

$$\dot{\pi}(t) = \pi(t)\Lambda.$$

- Given an initial $\pi(0)$ and a transition rate matrix Λ, we can calculate the probability of each state in any future t.
  - Often there is no analytical solution for this ODE.
  - However, in dynamic programming it is sufficient to only look at the *immediate future*.
  - The transition rates will be sufficient to do recursive optimization.

The stationary distribution

- A stationary distribution π^* satisfies

$$\pi^* \Lambda = 0.$$

Example 1: Employment and unemployment

- Remember the transition rate matrix:

$$\Lambda = \begin{bmatrix} -\delta & \delta \\ \lambda & -\lambda \end{bmatrix}$$

- We are looking for π_0^* and $\pi_1^* = 1 - \pi_0^*$ such that $\pi^* \Lambda = 0$.
- That is,

$$-\delta \pi_0^* + \lambda(1 - \pi_0^*) = 0.$$

Example 1: Employment and unemployment

$$-\delta\pi_0^* + \lambda(1 - \pi_0^*) = 0.$$

- This gives us
$$\pi_0^* = \frac{\lambda}{\lambda + \delta}.$$
- The same as $\pi_0(\infty)$ (not a coincidence).

Example 1: Employment and unemployment

$$-\delta\pi_0^* + \lambda(1 - \pi_0^*) = 0.$$

- This gives us
$$\pi_0^* = \frac{\lambda}{\lambda + \delta}.$$
- The same as $\pi_0(\infty)$ (not a coincidence).
- The steady-state probability of employment
  - increases in the job finding rate
  - decreases in the firing rate

Example 3: A faulty email server

- New emails arrive with rate λ:
  - $n \to n+1$
- With arrival rate η, all emails are lost
  - $n \to 0$

Example 3: A faulty email server

- The transition rate matrix: $$\begin{bmatrix} -\lambda & \lambda & 0 & \cdots \\ \eta & -\lambda-\eta & \lambda & \cdots \\ \eta & 0 & -\lambda-\eta & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Example 3: A faulty email server

- What is the stationary distribution of this process?
- For all $n \geq 0$:
$$\lambda\pi_n^* - (\lambda + \eta)\pi_{n+1}^* = 0$$
- This defines π_{n+1}^* recursively as
$$\pi_{n+1}^* = \frac{\lambda}{\lambda + \eta}\pi_n^*.$$
(A geometric distribution.)
- In turn,
$$\pi_0^* = \frac{\eta}{\lambda + \eta}.$$
(To make sure that π_n^*s sum to 1.)

A geometric distribution

The Poisson process

The Poisson process

- The possible states are $n = 0, 1, 2, ...$
- The Poisson process is characterized by an *arrival rate* λ (aka hazard rate).
- The transition rate matrix is

$$\begin{bmatrix} -\lambda & \lambda & 0 & \cdots \\ 0 & -\lambda & \lambda & \cdots \\ 0 & 0 & -\lambda & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

The Poisson process

- The possible states are $n = 0, 1, 2, ...$
- The Poisson process is characterized by an *arrival rate* λ (aka hazard rate).
- The transition rate matrix is

$$\begin{bmatrix} -\lambda & \lambda & 0 & \cdots \\ 0 & -\lambda & \lambda & \cdots \\ 0 & 0 & -\lambda & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

- The Poisson process is used to characterize *rare, memoryless* events.

Examples

- Phone calls to emergency center.
- Particles emitted via radioactive decay.
- Views of the CEU website.

Characterizing the Poisson process

The two key characteristics of the Poisson process

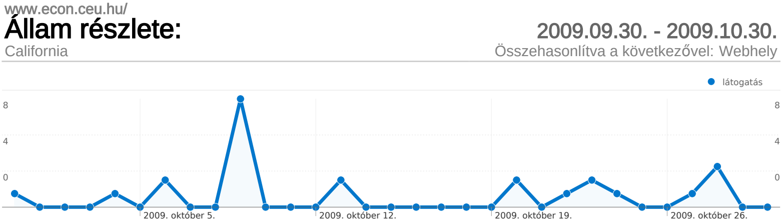
1. No two events happen at the same time ("rare events").
2. The future arrival of events is independent of past events ("memoryless").

Characterizing the Poisson process

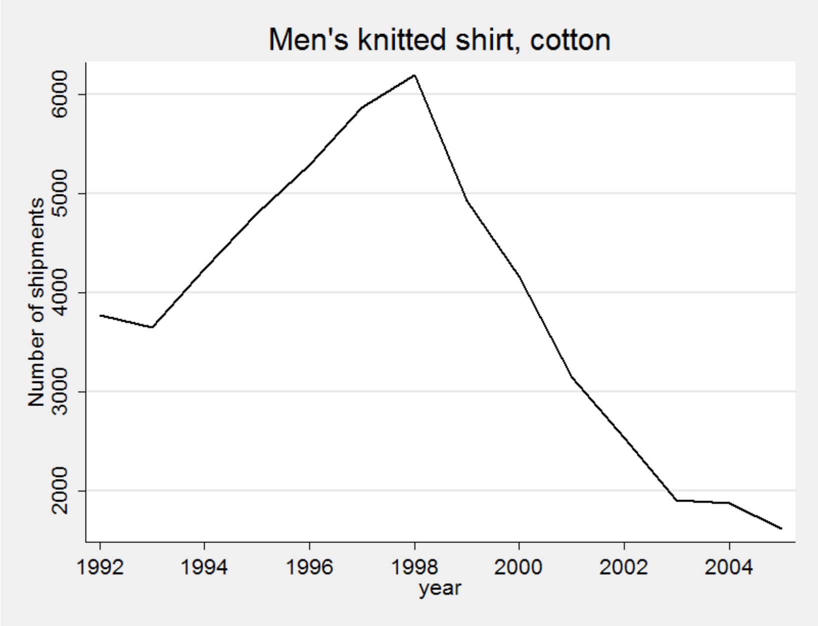
The Poisson process may arise

- from a truly memoryless process
  - radioactive decay
- from the law of small numbers
  - view of the CEU website from California

Visits to econ.ceu.hu from California



Export shipments of shirts from the U.S.



Export shipments of shirts from the U.S. to Iceland

Counterexamples

- Emergency phone calls during a natural disaster.
- Arrival of guests at a restaurant.
- Your phone calls to your mother.

Properties of the Poisson process

- The waiting time between the $n-1$sth and nth arrival is T_n.
- T_n is random, exponentially distributed with parameter λ:

$$\Pr(T_n \leq t) = 1 - \exp(-\lambda t).$$

- Waiting times are independent.

Properties of the Poisson process

- Let $N = n(t+h) - n(t)$ denote the number of arrivals between t and $t+h$.
- N is a Poisson-distributed random variable with parameter λh.
- It takes on values $0, 1, 2, \ldots$ with pdf

$$\Pr(n = k) = \frac{\exp(-\lambda h)(\lambda h)^k}{k!}$$

Properties of Poisson processes (continued)

- Take two independent Poisson processes with arrival λ_1 and λ_2.
  - The sum is also a Poisson process with arrival $\lambda_1 + \lambda_2$.
  - The waiting time for the first arrival is exponential with parameter $\lambda_1 + \lambda_2$.

Properties of Poisson processes (continued)

- Take two independent Poisson processes with arrival λ_1 and λ_2.
  - The sum is also a Poisson process with arrival $\lambda_1 + \lambda_2$.
  - The waiting time for the first arrival is exponential with parameter $\lambda_1 + \lambda_2$.
- Take a Poisson processes with arrival λ and a probability p.
- Kill each arrival with probability $1 - p$.
  - The new process is Poisson with arrival $p\lambda$.

Poisson representation of Markov chains

- Think of a Markov chain with N states.
- Starting in any given state, only $N-1$ things can happen (or nothing).
- Each $N-1$ jump has its own arrival rate.
- The first jump occurs with a Poisson arrival $\lambda_1 + ... + \lambda_{n-1}$ (see above).

Poisson representation of Markov chains (continued)

- Once there *is* a jump, which one is it?
- It could be any one of the $1, ..., n-1$.
- The probability of jump 1 is

$$\frac{\lambda_1}{\lambda_1 + ... + \lambda_{n-1}}.$$

Poisson representation of Markov chains (continued)

- Once there *is* a jump, which one is it?
- It could be any one of the $1, ..., n-1$.
- The probability of jump 1 is

$$\frac{\lambda_1}{\lambda_1 + ... + \lambda_{n-1}}.$$

- This is a good old probability $\in [0, 1]$.
- This looks more like a discrete transition matrix.

Poisson representation of Markov chains (continued)

- Once there *is* a jump, which one is it?
- It could be any one of the $1, ..., n-1$.
- The probability of jump 1 is

$$\frac{\lambda_1}{\lambda_1 + ... + \lambda_{n-1}}.$$

- This is a good old probability $\in [0, 1]$.
- This looks more like a discrete transition matrix.
- I find it useful to think about Markov chains as the sum of Poisson processes.

Checklist

By now you should understand

1. continuous-time Markov chain
2. arrival rate matrix
3. forward Kolmogorov equation
4. stationary distribution
5. Poisson process
6. Poisson distribution

Appendix

Big-O, small-o

Big-O

A function $f(x)$ is $O(g(x))$ for a known function $g(x)$ if

$$\lim_{x \to 0} \frac{f(x)}{g(x)} < \infty$$

Small-o

A function $f(x)$ is $o(g(x))$ for a known function $g(x)$ if

$$\lim_{x \to 0} \frac{f(x)}{g(x)} = 0$$

Examples

- $f(x) = x^2$ is both $O(x)$ and $o(x)$. It is also $O(x^2)$.
- $f(x) = x^2 + 2x$ is $O(x)$ but not $o(x)$.
- $f(x) = x^2 + 2x + 4$ is not $O(x)$.