Model Checking Markov Chains

Lecture 4: Continuous-Time Markov Chains

Joost-Pieter Katoen

Software Modeling and Verification Group

RWTH Aachen University

affiliated to University of Twente, Formal Methods and Tools

Lecture at Model Checking Summerschool, October 12, 2010

Content of this lecture

$\Rightarrow$ Negative exponential distribution

- definition, usage, properties

- Continuous-time Markov chains
  - definition, semantics, examples
- Performance measures
  - transient and steady-state probabilities, uniformization

Time in DTMCs

- Time in a DTMC proceeds in discrete steps
- Two possible interpretations
  - accurate model of (discrete) time units
    - e.g., clock ticks in model of an embedded device
  - time-abstract
    - no information assumed about the time transitions take
- Continuous-time Markov chains (CTMCs)
  - dense model of time
  - transitions can occur at any (real-valued) time instant
  - modelled using negative exponential distributions

Continuous random variables

- X is a random variable (r.v., for short)
  - on a sample space with probability measure Pr
  - assume the set of possible values that X may take is dense

- X is *continuously distributed* if there exists a function $f(x)$ such that:

$$\mathrm{Pr}\{X \leqslant d\} = \int_{-\infty}^{d} f(x)\, dx \quad \text{for each real number } d$$

where f satisfies: $f(x) \geqslant 0$ for all x and $\int_{-\infty}^{\infty} f(x)\, dx = 1$

  - $F_X(d) = \mathrm{Pr}\{X \leqslant d\}$ is the *(cumulative) probability distribution function*
  - $f(x)$ is the *probability density function*

Negative exponential distribution

The density of an *exponentially distributed* r.v. Y with rate $\lambda \in \mathbb{R}_{>0}$ is:

$$f_Y(x) = \lambda \cdot e^{-\lambda \cdot x} \quad \text{for } x > 0 \quad \text{and } f_Y(x) = 0 \text{ otherwise}$$

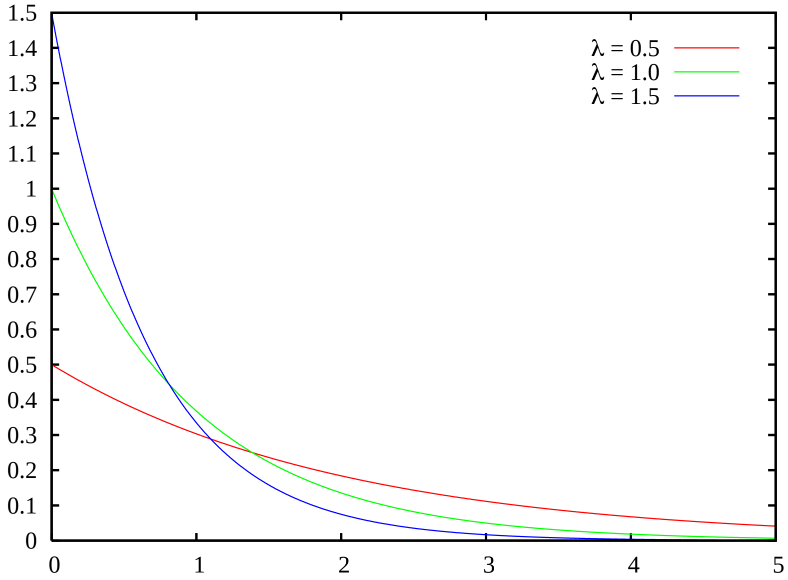
The cumulative distribution of Y:

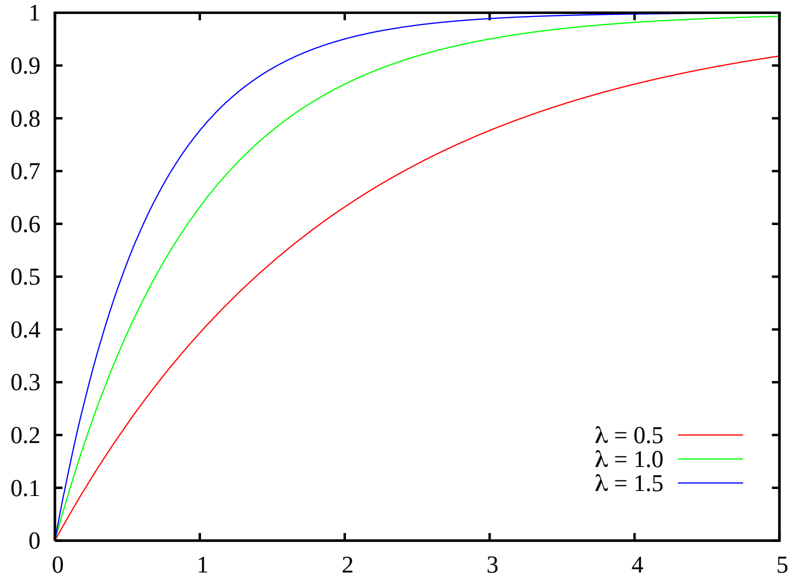
$$F_Y(d) \;=\; \int_0^d \lambda \cdot e^{-\lambda \cdot x}\, dx \;=\; \left[-e^{-\lambda \cdot x}\right]_0^d \;=\; 1 - e^{-\lambda \cdot d}$$

- expectation $E[Y] = \int_0^\infty x \cdot \lambda \cdot e^{-\lambda \cdot x}\, dx = \frac{1}{\lambda}$
- variance $\mathit{Var}[Y] = \frac{1}{\lambda^2}$

the rate $\lambda \in \mathbb{R}_{>0}$ uniquely determines an exponential distribution.

Exponential pdf and cdf





the higher λ, the faster the cdf approaches 1

Why exponential distributions?

- Are *adequate* for many real-life phenomena
  - the time until a radioactive particle decays
  - the time between successive car accidents
  - inter-arrival times of jobs, telephone calls in a fixed interval
- Are the continuous counterpart of *geometric* distribution
- Heavily used in physics, performance, and reliability analysis
- Can *approximate* general distributions arbitrarily closely
- Yield a *maximal entropy* if only the mean is known

Memoryless property

1. For any random variable X with an exponential distribution:

$$\Pr\{X > t + d \mid X > t\} \; = \; \Pr\{X > d\} \text{ for any } t, d \in \mathbb{R}_{\geqslant 0}.$$

2. Any continuous distribution which is memoryless is an exponential one.

Proof of 1. : Let λ be the rate of X's distribution. Then we derive:

$$\Pr\{X > t + d \mid X > t\} \; = \; \frac{\Pr\{X > t{+}d \cap X > t\}}{\Pr\{X > t\}} \; = \; \frac{\Pr\{X > t{+}d\}}{\Pr\{X > t\}}$$

$$= \; \frac{e^{-\lambda \cdot (t+d)}}{e^{-\lambda \cdot t}} \; = \; e^{-\lambda \cdot d} \; = \; \Pr\{X > d\}.$$

Proof of 2. : by contradiction, using the total law of probability.

Closure under minimum

For independent, exponentially distributed random variables X and Y with rates $\lambda, \mu \in \mathbb{R}_{>0}$, r.v. $\min(X, Y)$ is exponentially distributed with rate $\lambda+\mu$, i.e.,:

$$\Pr\{\min(X, Y) \leqslant t\} \;=\; 1 - e^{-(\lambda+\mu)\cdot t} \quad \text{for all } t \in \mathbb{R}_{\geqslant 0}$$

Proof

Let λ (μ) be the rate of X's (Y's) distribution. Then we derive:

$$\Pr\{\min(X, Y) \leqslant t\} = \Pr_{X,Y}\{(x, y) \in \mathbb{R}^2_{\geqslant 0} \mid \min(x, y) \leqslant t\}$$

$$= \int_0^\infty \left(\int_0^\infty \mathbf{I}_{\min(x,y)\leqslant t}(x, y) \cdot \lambda e^{-\lambda x} \cdot \mu e^{-\mu y} \, dy \right) dx$$

$$= \int_0^t \int_x^\infty \lambda e^{-\lambda x} \cdot \mu e^{-\mu y} \, dy \, dx + \int_0^t \int_y^\infty \lambda e^{-\lambda x} \cdot \mu e^{-\mu y} \, dx \, dy$$

$$= \int_0^t \lambda e^{-\lambda x} \cdot e^{-\mu x} \, dx + \int_0^t e^{-\lambda y} \cdot \mu e^{-\mu y} \, dy$$

$$= \int_0^t \lambda e^{-(\lambda+\mu)x} \, dx + \int_0^t \mu e^{-(\lambda+\mu)y} \, dy$$

$$= \int_0^t (\lambda+\mu) \cdot e^{-(\lambda+\mu)z} \, dz \; = \; 1 - e^{-(\lambda+\mu)t}$$

Winning the race with two competitors

For independent, exponentially distributed random variables

X and Y with rates $\lambda, \mu \in \mathbb{R}_{>0}$, it holds:

$$\Pr\{X \leqslant Y\} = \frac{\lambda}{\lambda + \mu}$$

Proof

Let λ (μ) be the rate of X's (Y's) distribution. Then we derive:

$$
\begin{aligned}
\Pr\{X \leqslant Y\} &= \Pr_{X,Y}\{(x,y) \in \mathbb{R}^2_{\geqslant 0} \mid x \leqslant y\} \\
&= \int_0^\infty \mu e^{-\mu y} \left(\int_0^y \lambda e^{-\lambda x}\, dx \right) dy \\
&= \int_0^\infty \mu e^{-\mu y} \left(1 - e^{-\lambda y} \right) dy \\
&= 1 - \int_0^\infty \mu e^{-\mu y} \cdot e^{-\lambda y}\, dy \;=\; 1 - \int_0^\infty \mu e^{-(\mu+\lambda)y}\, dy \\
&= 1 - \frac{\mu}{\mu+\lambda} \cdot \underbrace{\int_0^\infty (\mu+\lambda) e^{-(\mu+\lambda)y}\, dy}_{=1} \\
&= 1 - \frac{\mu}{\mu+\lambda} \;=\; \frac{\lambda}{\mu+\lambda}
\end{aligned}
$$

Winning the race with many competitors

For independent, exponentially distributed random variables

$X_1, X_2, \ldots, X_n$ with rates $\lambda_1, \ldots, \lambda_n \in \mathbb{R}_{>0}$, it holds:

$$\Pr\{X_i = \min(X_1, \ldots, X_n)\} = \frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j}$$

Content of this lecture

- Negative exponential distribution
  - definition, usage, properties

$\Rightarrow$ Continuous-time Markov chains
  - definition, semantics, examples

- Performance measures
  - transient and steady-state probabilities, uniformization

Continuous-time Markov chain

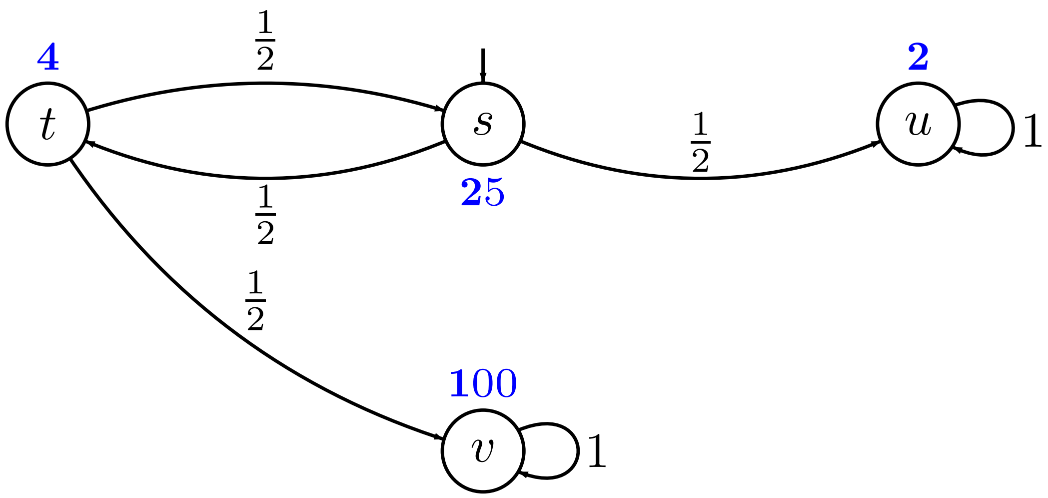
A *continuous-time Markov chain* (CTMC) is a tuple $(S, \mathbf{P}, r, L)$ where:

- S is a countable (today: finite) set of *states*
- $\mathbf{P} : S \times S \rightarrow [0, 1]$, a *stochastic matrix*
  - $\mathbf{P}(s, s')$ is one-step probability of going from state s to state s'
  - s is called *absorbing* iff $\mathbf{P}(s, s) = 1$
- $r : S \rightarrow \mathbb{R}_{>0}$, the *exit-rate function*
  - $r(s)$ is the rate of exponential distribution of residence time in state s

$\Rightarrow$ a CTMC is a Kripke structure with random state residence times

Continuous-time Markov chain

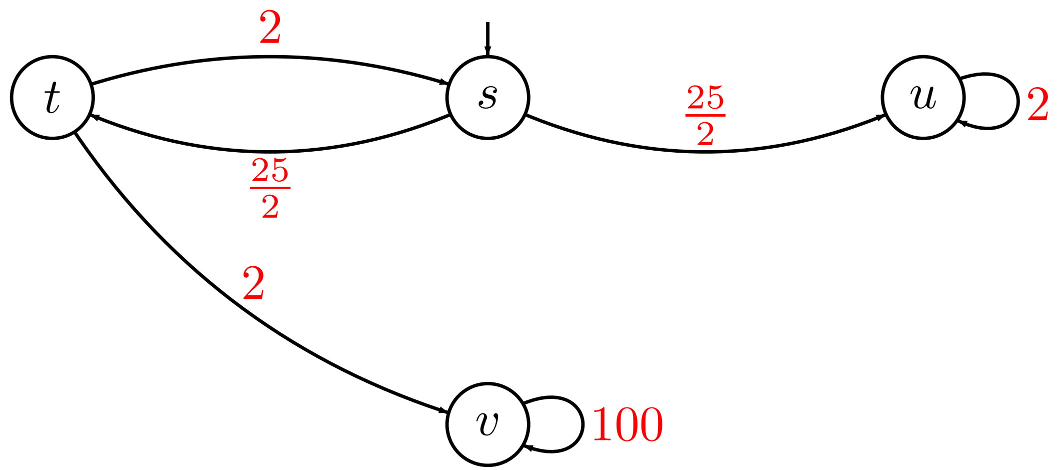
a CTMC $(S, \mathbf{P}, r, L)$ is a DTMC plus an exit-rate function $r : S \rightarrow \mathbb{R}_{>0}$



the average residence time in state s is $\dfrac{1}{r(s)}$

A classical (though equivalent) perspective

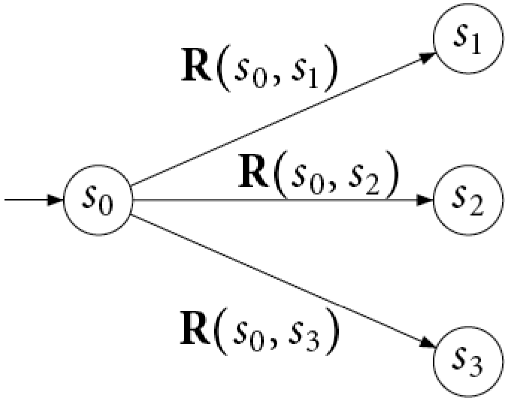
a CTMC is a triple $(S, \mathbf{R}, L)$ with $\mathbf{R}(s, s') = \mathbf{P}(s, s') \cdot r(s)$



CTMC semantics: example

- Transition $s \longrightarrow s'$:= r.v. $X_{s,s'}$ with rate $\mathbf{R}(s, s')$

- Probability to go from state s_0 to, say, state s_2 is:



$$\Pr\{X_{s_0,s_2} \leqslant X_{s_0,s_1} \cap X_{s_0,s_2} \leqslant X_{s_0,s_3}\}$$

$$=$$

$$\frac{\mathbf{R}(s_0, s_2)}{\mathbf{R}(s_0, s_1) + \mathbf{R}(s_0, s_2) + \mathbf{R}(s_0, s_3)} = \frac{\mathbf{R}(s_0, s_2)}{r(s_0)}$$

- Probability of staying at most t time in s_0 is:

$$\Pr\{\min(X_{s_0,s_1}, X_{s_0,s_2}, X_{s_0,s_3}) \leqslant t\}$$

$$=$$

$$1 - e^{-(\mathbf{R}(s_0,s_1)+\mathbf{R}(s_0,s_2)+\mathbf{R}(s_0,s_3))\cdot t} = 1 - e^{-r(s_0)\cdot t}$$

CTMC semantics

- The probability that transition $s \longrightarrow s'$ is *enabled* in $[0, t]$:

$$1 - e^{-\mathbf{R}(s,s')\cdot t}$$

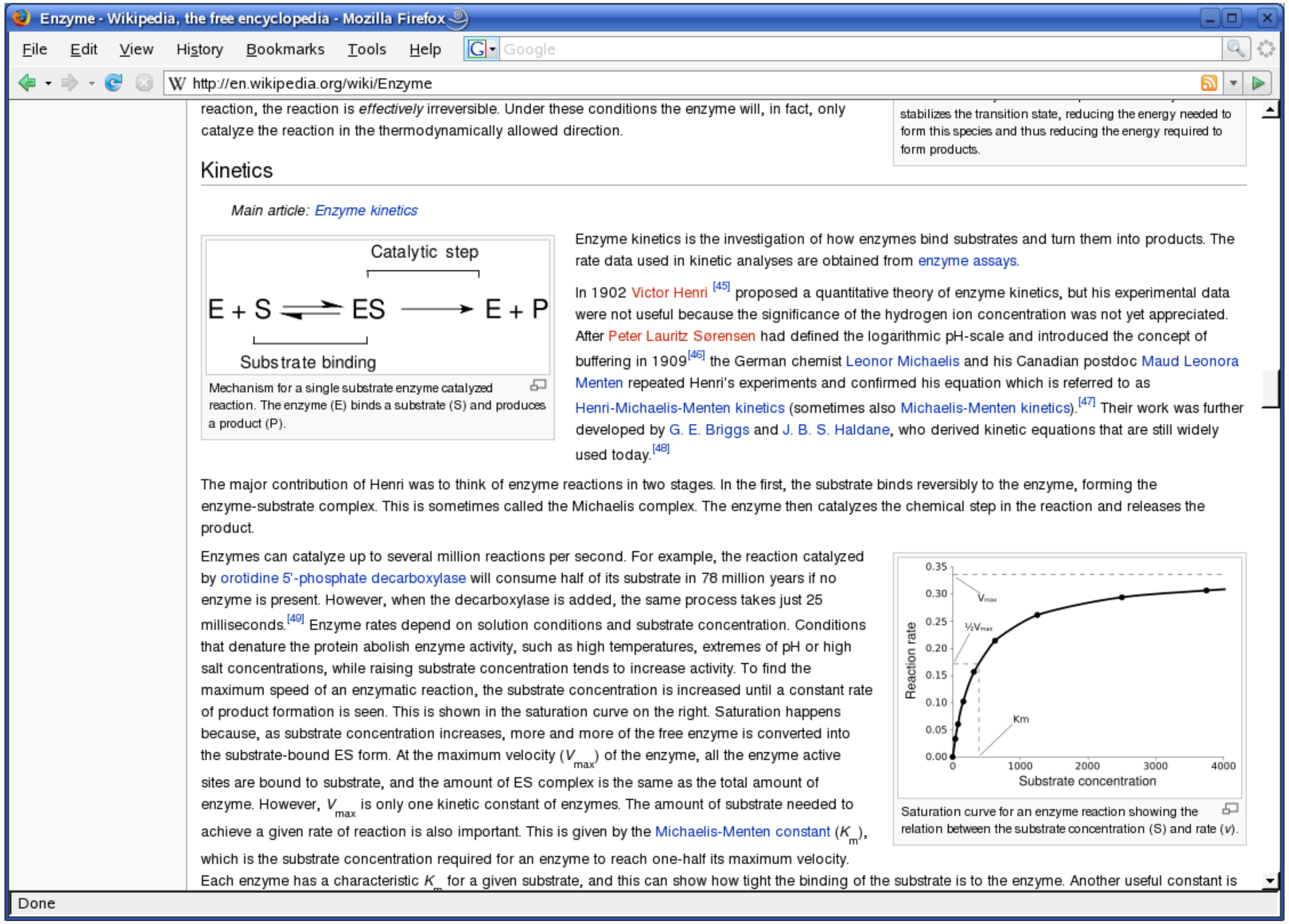
- The probability to *move* from non-absorbing s to s' in $[0, t]$ is:

$$\frac{\mathbf{R}(s, s')}{r(s)} \cdot \left(1 - e^{-r(s)\cdot t}\right)$$

- The probability to *take some* outgoing transition from s in $[0, t]$ is:

$$\int_0^t r(s)\cdot e^{-r(s)\cdot x}\, dx \;=\; 1 - e^{-r(s)\cdot t}$$

Enzyme-catalysed substrate conversion



Enzyme - Wikipedia, the free encyclopedia - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://en.wikipedia.org/wiki/Enzyme

reaction, the reaction is *effectively* irreversible. Under these conditions the enzyme will, in fact, only catalyze the reaction in the thermodynamically allowed direction.

stabilizes the transition state, reducing the energy needed to form this species and thus reducing the energy required to form products.

Kinetics

Main article: Enzyme kinetics

Mechanism for a single substrate enzyme catalyzed reaction. The enzyme (E) binds a substrate (S) and produces a product (P).

Enzyme kinetics is the investigation of how enzymes bind substrates and turn them into products. The rate data used in kinetic analyses are obtained from enzyme assays.

In 1902 Victor Henri [45] proposed a quantitative theory of enzyme kinetics, but his experimental data were not useful because the significance of the hydrogen ion concentration was not yet appreciated. After Peter Lauritz Sørensen had defined the logarithmic pH-scale and introduced the concept of buffering in 1909[46] the German chemist Leonor Michaelis and his Canadian postdoc Maud Leonora Menten repeated Henri's experiments and confirmed his equation which is referred to as Henri-Michaelis-Menten kinetics (sometimes also Michaelis-Menten kinetics).[47] Their work was further developed by G. E. Briggs and J. B. S. Haldane, who derived kinetic equations that are still widely used today.[48]

The major contribution of Henri was to think of enzyme reactions in two stages. In the first, the substrate binds reversibly to the enzyme, forming the enzyme-substrate complex. This is sometimes called the Michaelis complex. The enzyme then catalyzes the chemical step in the reaction and releases the product.

Enzymes can catalyze up to several million reactions per second. For example, the reaction catalyzed by orotidine 5'-phosphate decarboxylase will consume half of its substrate in 78 million years if no enzyme is present. However, when the decarboxylase is added, the same process takes just 25 milliseconds.[49] Enzyme rates depend on solution conditions and substrate concentration. Conditions that denature the protein abolish enzyme activity, such as high temperatures, extremes of pH or high salt concentrations, while raising substrate concentration tends to increase activity. To find the maximum speed of an enzymatic reaction, the substrate concentration is increased until a constant rate of product formation is seen. This is shown in the saturation curve on the right. Saturation happens because, as substrate concentration increases, more and more of the free enzyme is converted into the substrate-bound ES form. At the maximum velocity (V_{max}) of the enzyme, all the enzyme active sites are bound to substrate, and the amount of ES complex is the same as the total amount of enzyme. However, V_{max} is only one kinetic constant of enzymes. The amount of substrate needed to achieve a given rate of reaction is also important. This is given by the Michaelis-Menten constant (K_m), which is the substrate concentration required for an enzyme to reach one-half its maximum velocity. Each enzyme has a characteristic K_m for a given substrate, and this can show how tight the binding of the substrate is to the enzyme. Another useful constant is

Saturation curve for an enzyme reaction showing the relation between the substrate concentration (S) and rate (v).

Done

Stochastic chemical kinetics

- Types of reaction described by stochiometric equations:

$$E + S \underset{k_2}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_3} E + P$$

- N different types of molecules that randomly collide

  where state $X(t) = (x_1, \ldots, x_N)$ with $x_i = \#$ molecules of sort i

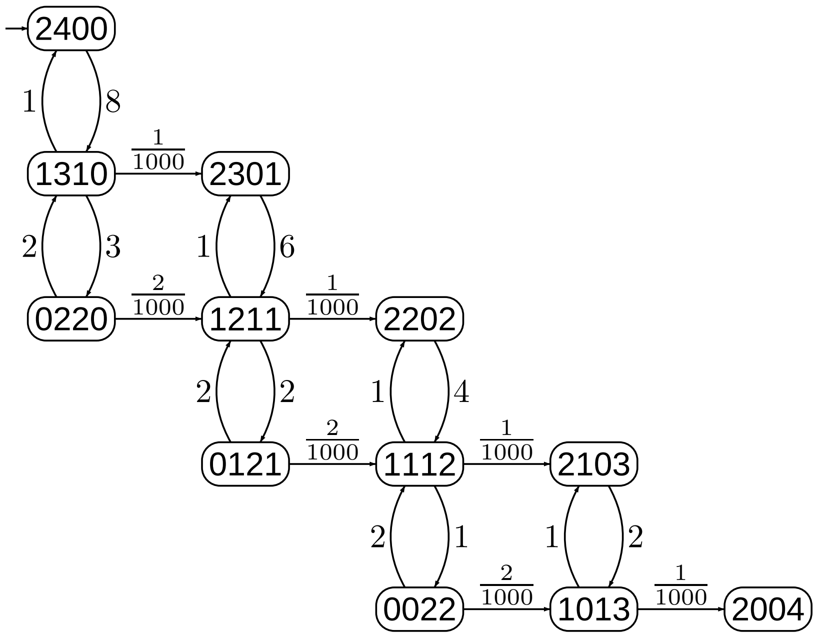
- Reaction probability within infinitesimal interval $[t, t{+}\Delta)$:

$$\alpha_m(\vec{x}) \cdot \Delta \;=\; \Pr\{\text{reaction } m \text{ in } [t, t{+}\Delta) \mid X(t) = \vec{x}\}$$

  where $\alpha_m(\vec{x}) = k_m \cdot \#$ possible combinations of reactant molecules in $\vec{x}$

- Process is a continuous-time Markov chain

Enzyme-catalyzed substrate conversion as a CTMC



States:	*init*	*goal*
enzymes	2	2
substrates	4	0
complex	0	0
products	0	4

Transitions: $E + S \underset{1}{\overset{1}{\rightleftharpoons}} C \xrightarrow{0.001} E + P$

e.g., $(x_E, x_S, x_C, x_P) \xrightarrow{0.001 \cdot x_C} (x_E + 1, x_S, x_C - 1, x_P + 1)$ for $x_C > 0$

CTMCs are omnipresent!

- Markovian queueing networks (Kleinrock 1975)
- Stochastic Petri nets (Molloy 1977)
- Stochastic activity networks (Meyer & Sanders 1985)
- Stochastic process algebra (Herzog *et al.*, Hillston 1993)
- Probabilistic input/output automata (Smolka *et al.* 1994)
- Calculi for biological systems (Priami *et al.*, Cardelli 2002)

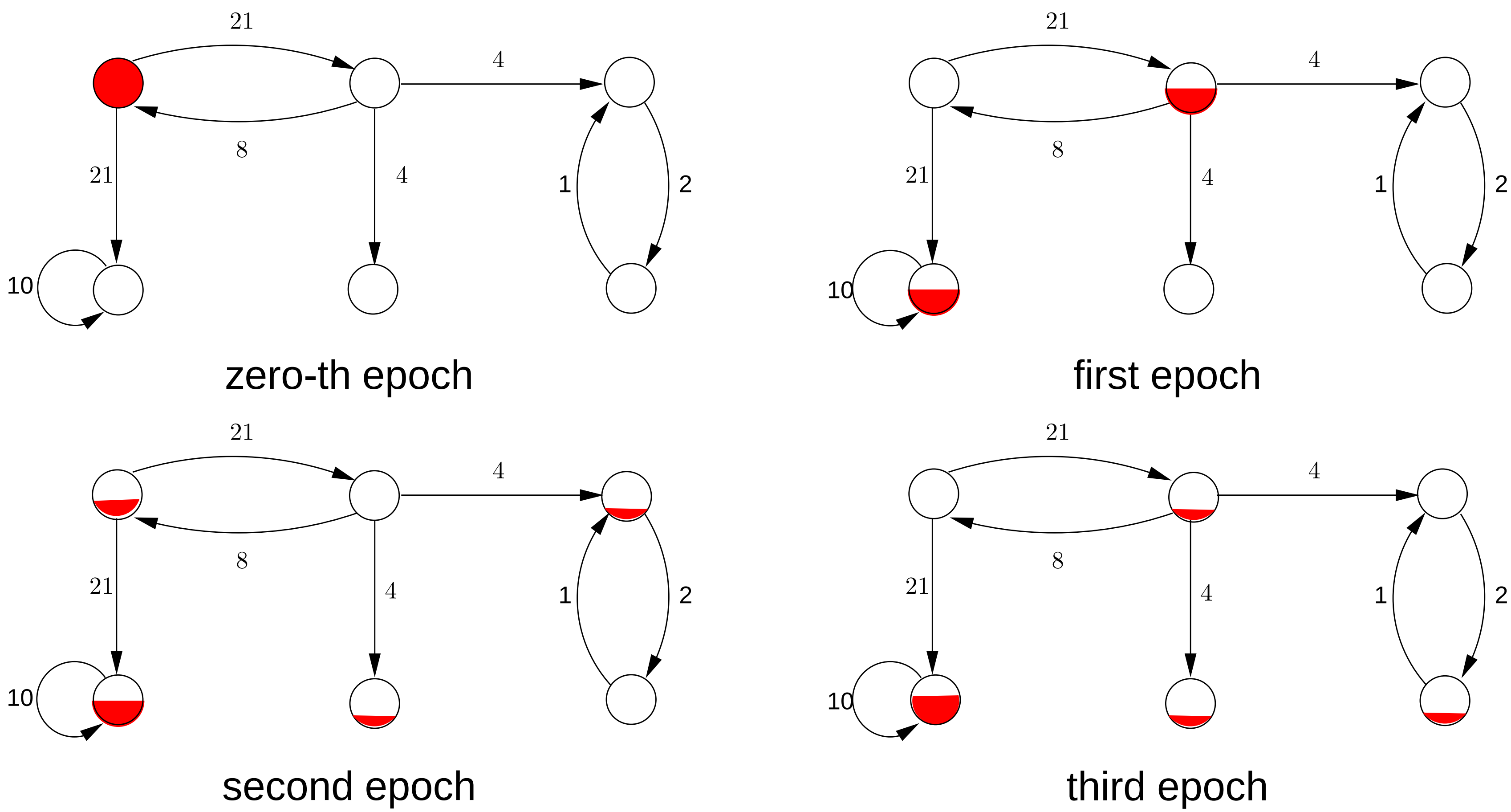
CTMCs are one of the most prominent models in performance analysis

Content of this lecture

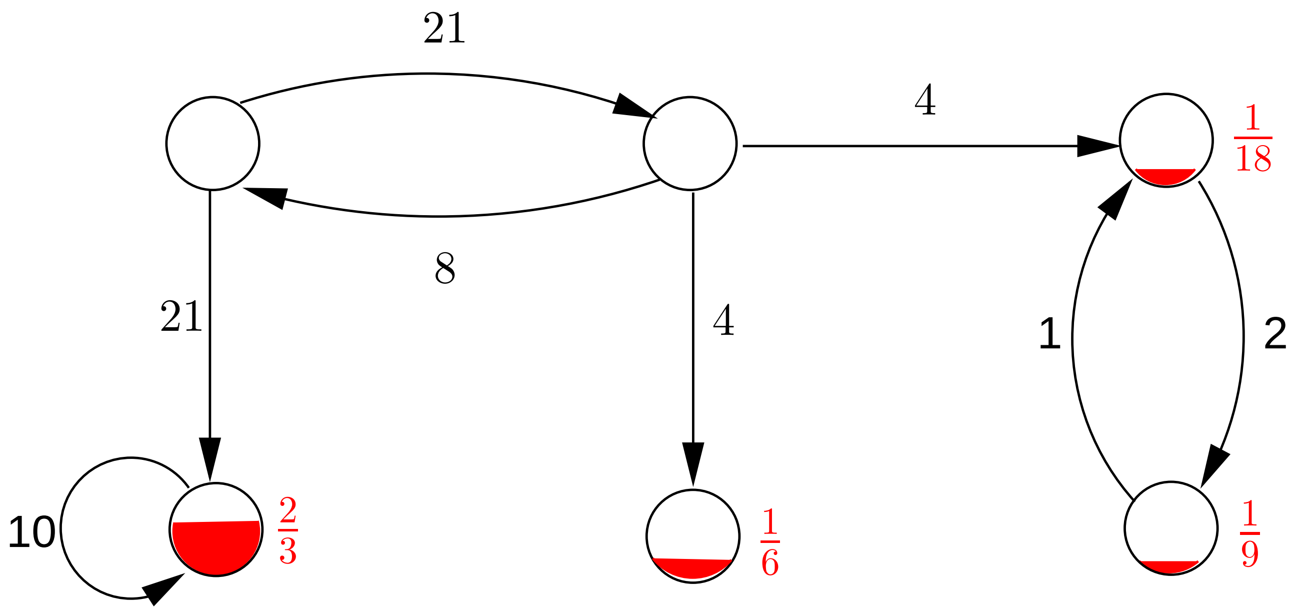
- Negative exponential distribution
  - definition, usage, properties
- Continuous-time Markov chains
  - definition, semantics, examples

$\Rightarrow$ Performance measures
  - transient and steady-state probabilities, uniformization

Time-abstract evolution of a CTMC



On the long run



Transient distribution of a CTMC

Let $X(t)$ denote the state of a CTMC at time $t \in \mathbb{R}_{\geqslant 0}$.

Probability to be in state s at time t:

$$\begin{aligned} p_s(t) &= \Pr\{ X(t) = s \} \\ &= \sum_{s' \in S} \Pr\{ X(0) = s' \} \cdot \Pr\{ X(t) = s \mid X(0) = s' \} \end{aligned}$$

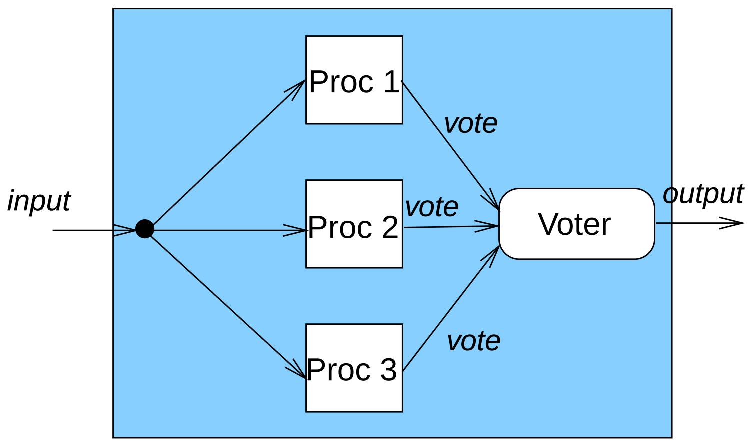
Transient probability vector $\underline{p}(t) = (p_{s_1}(t), \ldots, p_{s_k}(t))$ satisfies:

$$\underline{p}'(t) = \underline{p}(t) \cdot (\mathbf{R} - \mathbf{r}) \qquad \text{given} \qquad \underline{p}(0)$$

where $\mathbf{r}$ is the diagonal matrix of vector $\underline{r}$.

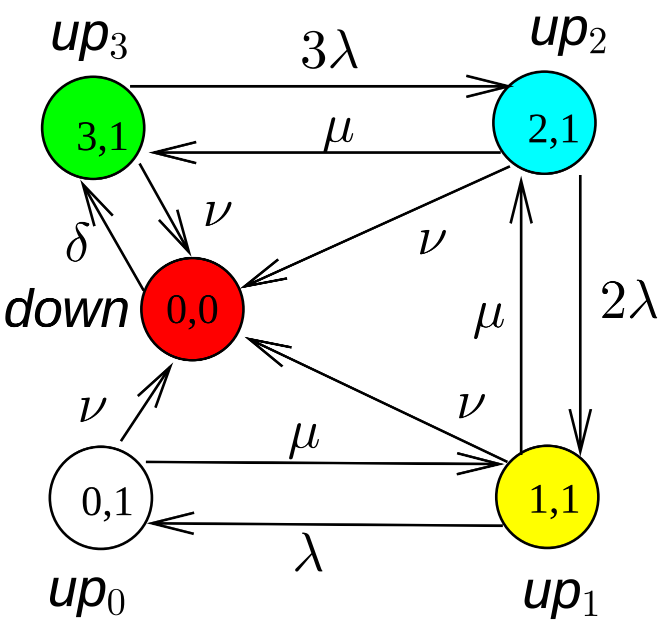
A triple modular redundant system

- 3 processors and a single voter:
  - processors run same program; voter takes a majority vote
  - each component (processor and voter) is failure-prone
  - there is a single repairman for repairing processors and voter



- Modelling assumptions:
  - if voter fails, entire system goes down
  - after voter-repair, system starts “as new”
  - state $=$ ($\#$processors, $\#$voters)

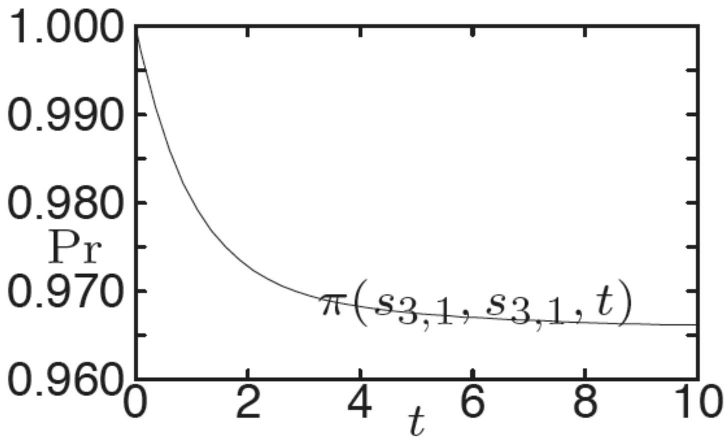
Modelling a TMR system as a CTMC



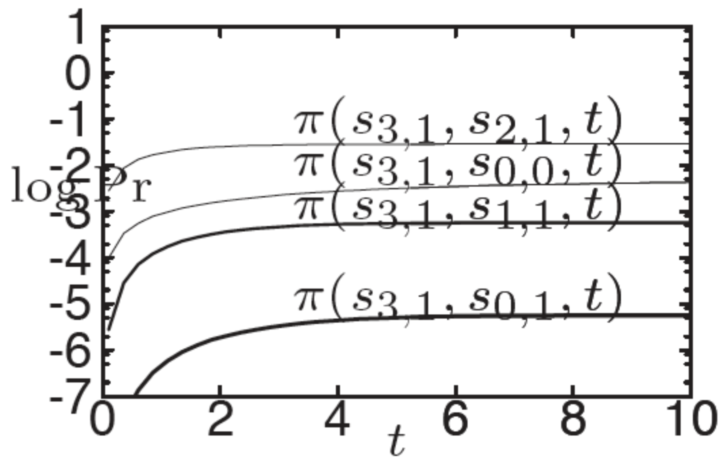
- processor failure rate is λ fph;
  its repair rate is μ rph
- voter failure rate is ν fph;
  its repair rate is δ rph
- rate matrix: e.g., $\mathbf{R}((3,1),(2,1)) = 3\lambda$
- exit rates: e.g., $r((3,1)) = 3\lambda + \nu$
- probability matrix: e.g.,

$$\mathbf{P}((3,1),(2,1)) = \frac{3\lambda}{3\lambda + \nu}$$

Transient probabilities



$p_{s_{3,1}}(t)$ for $t \leqslant$ 10 hours



$p(t)$ for $t \leqslant 10$ hours (log-scale)

$\lambda = 0.01$ fph, $\nu = 0.001$ fph

$\mu = 1$ rph and $\delta = 0.2$ rph

(© book by B.R. Haverkort)

Steady-state distribution of a CTMC

For any finite and strongly connected CTMC it holds:

$$p_s = \lim_{t \to \infty} p_s(t) \quad \Leftrightarrow \quad \lim_{t \to \infty} p'_s(t) = 0 \quad \Leftrightarrow \quad \lim_{t \to \infty} p_s(t) \cdot (\mathbf{R} - \mathbf{r}) = 0$$

Steady-state probability vector $\underline{p} = (p_{s_1}, \ldots, p_{s_k})$ satisfies:

$$\underline{p} \cdot (\mathbf{R} - \mathbf{r}) \; = \; 0 \quad \text{where} \quad \sum_{s \in S} p_s = 1$$

Steady-state distribution

s	$s_{3,1}$	$s_{2,1}$	$s_{1,1}$	$s_{0,1}$	$s_{0,0}$
$p(s)$	$9.655 \cdot 10^{-1}$	$2.893 \cdot 10^{-2}$	$5.781 \cdot 10^{-4}$	$5.775 \cdot 10^{-6}$	$4.975 \cdot 10^{-3}$

The probability of $\geqslant$ two processors and the voter are up

once the CTMC has reached an equilibrium is 0.9655+0.02893 $\approx$ 0.993

$\lambda = 0.01$ fph, $\nu = 0.001$ fph

$\mu = 1$ rph and $\delta = 0.2$ rph

Computing transient probabilities

- Transient probability vector $\underline{p}(t) = (p_{s_1}(t), \ldots, p_{s_k}(t))$ satisfies:

$$\underline{p}'(t) = \underline{p}(t) \cdot (\mathbf{R}-\mathbf{r}) \quad \text{given} \quad \underline{p}(0)$$

- Solution using Taylor-Maclaurin expansion:

$$\underline{p}(t) \; = \; \underline{p}(0) \cdot e^{(\mathbf{R}-\mathbf{r}) \cdot t} \; = \; \underline{p}(0) \cdot \sum_{i=0}^{\infty} \frac{((\mathbf{R}-\mathbf{r}) \cdot t)^i}{i!}$$

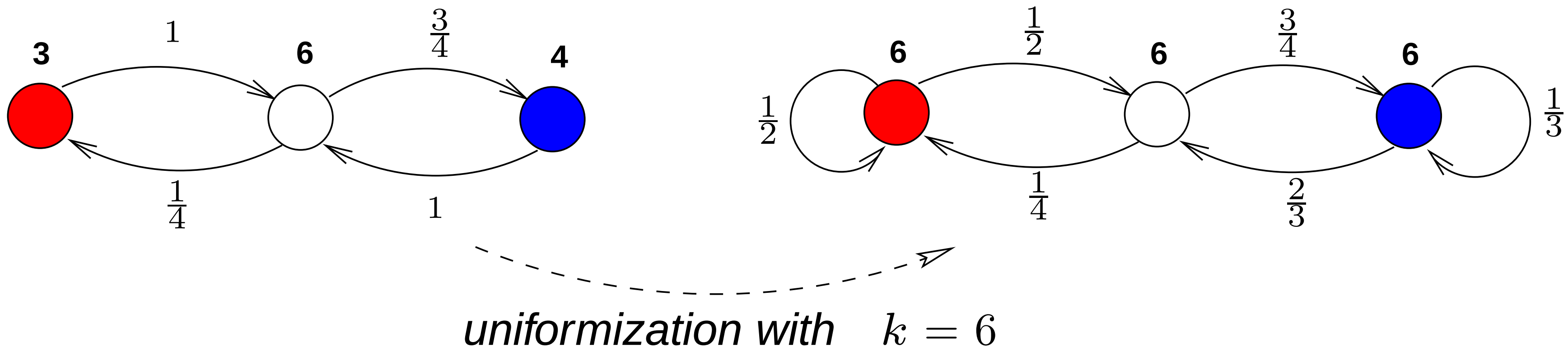
- Main problems: infinite summation + numerical instability due to
  - non-sparsity of $(\mathbf{R}-\mathbf{r})^i$ and presence positive and negative entries

Uniform CTMCs

- A CTMC is uniform if $r(s) = r$ for all s for some $r \in \mathbb{R}_{>0}$
- Any CTMC can be changed into a weak bisimilar uniform CTMC
- Let $r \in \mathbb{R}_{>0}$ such that $r \geqslant \mathit{max}_{s \in S}\ r(s)$
  - $\frac{1}{r}$ is at most the shortest mean residence time in CTMC $\mathcal{C}$
- Then $u(r, \mathcal{C}) = (S, \overline{\mathbf{P}}, \overline{r}, L)$ with $\overline{r}(s) = r$ for any s, and:

$$\overline{\mathbf{P}}(s, s') = \frac{r(s)}{r} \cdot \mathbf{P}(s, s') \text{ if } s' \neq s \quad \text{and} \quad \overline{\mathbf{P}}(s, s) = \frac{r(s)}{r} \cdot \mathbf{P}(s, s) + 1 - \frac{r(s)}{r}$$

Uniformization



all state transitions in CTMC $u(r, \mathcal{C})$ occur at an average pace of r per time unit

Computing transient probabilities

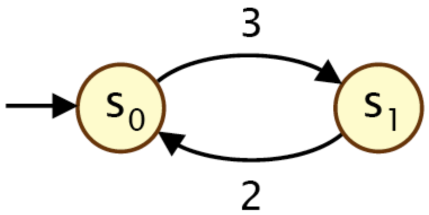
- Now: $\underline{p}(t) = \underline{p}(0) \cdot e^{r \cdot (\overline{\mathbf{P}} - \mathbf{I})t} = \underline{p}(0) \cdot e^{-rt} \cdot e^{r \cdot t \cdot \overline{\mathbf{P}}} = \underline{p}(0) \cdot \sum_{i=0}^{\infty} \underbrace{e^{-r \cdot t} \frac{(r \cdot t)^i}{i!}}_{\text{Poisson prob.}} \cdot \overline{\mathbf{P}}^i$

- Summation can be truncated *a priori* for a given error bound $\varepsilon > 0$:

$$\left\| \sum_{i=0}^{\infty} e^{-rt} \frac{(rt)^i}{i!} \cdot \underline{p}(i) - \sum_{i=0}^{k_\varepsilon} e^{-rt} \frac{(rt)^i}{i!} \cdot \underline{p}(i) \right\| = \left\| \sum_{i=k_\varepsilon+1}^{\infty} e^{-rt} \frac{(rt)^i}{i!} \cdot \underline{p}(i) \right\|$$

- Choose k_ε minimal s.t.: $\sum_{i=k_{\varepsilon+1}}^{\infty} e^{-rt} \frac{(rt)^i}{i!} = 1 - \sum_{i=0}^{k_\varepsilon} e^{-rt} \frac{(rt)^i}{i!} \leqslant \varepsilon$

Transient probabilities: example



$\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \underline{r} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ and $\overline{\mathbf{P}}_3 = \begin{bmatrix} 0 & 1 \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}$

Let initial distribution $\underline{p}(0) = (1, 0)$, and time bound $t{=}1$.

Then:

$$\underline{p}(0) \cdot \sum_{i=0}^{\infty} e^{-3} \frac{3^i}{i!} \cdot \overline{\mathbf{P}}^i$$

$$= (1,0) \cdot e^{-3} \tfrac{1}{0!} \cdot \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + (1,0) \cdot e^{-3} \tfrac{3}{1!} \cdot \begin{bmatrix} 0 & 1 \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

$$+ (1,0) \cdot e^{-3} \tfrac{9}{2!} \cdot \begin{bmatrix} 0 & 1 \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}^2 + \ldots\ldots$$

$$\approx (0.404043, 0.595957)$$

CTMC paths

- An infinite path σ in a CTMC $\mathcal{C} = (S, \mathbf{P}, r, L)$ is of the form:

$$\sigma \;=\; s_0 \xrightarrow{t_0} s_1 \xrightarrow{t_1} s_2 \xrightarrow{t_2} s_3 \ldots\ldots$$

  with s_i is a state in S, $t_i \in \mathbb{R}_{>0}$ is a duration, and $\mathbf{P}(s_i, s_{i+1}) > 0$.

- A Borel space on infinite paths exists (cylinder construction)
  - reachability, timed reachability, and ω-regular properties are measurable

- A path is Zeno if $\sum_i t_i$ is converging

- Theorem: the probability of the set of Zeno paths in any CTMC is 0

Summarizing

- Negative exponential distribution
  - suitable for many practical phenomena
  - nice mathematical properties
- Continuous-time Markov chains
  - Kripke structures with exponential state residence times
  - used in many different fields, e.g., performance, biology, . . .
- Performance measures
  - transient probability vector: where is a CTMC at time t?
  - steady-state probability vector: where is a CTMC on the long run?

谢谢大家！