Lumping continuous-time Markov chains

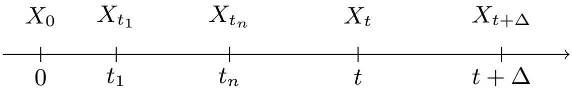
Alexander Erreygers Jasper De Bock

Ghent University

WUML & WPMSIIP 2017

We are interested in making inferences about the *state* of some *system*.

The state at time $t \in \mathbb{R}_{\geq 0}$ is denoted by X_t, and takes values in a finite *state space* $\mathscr{X}$.

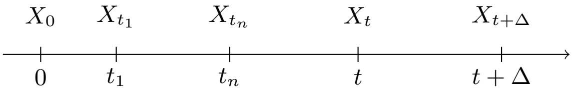


Precise continuous-time Markov chains

General stochastic processes

We are interested in making inferences about the *state* of some *system*.

The state at time $t \in \mathbb{R}_{\geq 0}$ is denoted by X_t, and takes values in a finite *state space* $\mathscr{X}$.



To make inferences, we need to specify probabilities of the form

$$\mathrm{P}(X_0 = x)$$

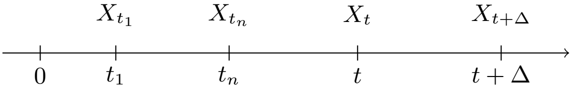
and

$$\mathrm{P}(X_{t+\Delta} = y | X_{t_1} = x_1, \ldots, X_{t_n} = x_n, X_t = x),$$

where $x_1, \ldots, x_n, x, y \in \mathscr{X}$, $0 \leq t_1 < \cdots < t_n < t$ and $\Delta \in \mathbb{R}_{\geq 0}$.

Precise continuous-time Markov chains

The Markov property



A stochastic process is called a *continuous-time Markov chain* (CTMC) if (our beliefs about) the future only depends on (our knowledge of) the present and not on the past:

$$\mathrm{P}(X_{t+\Delta} = y | X_{t_1} = x_1, \ldots, X_{t_n} = x_n, X_t = x) = \mathrm{P}(X_{t+\Delta} = y | X_t = x).$$

Precise continuous-time Markov chains

The Markov property

A stochastic process is called a *continuous-time Markov chain* (CTMC) if (our beliefs about) the future only depends on (our knowledge of) the present and not on the past:

$$\mathrm{P}(X_{t+\Delta} = y | X_{t_1} = x_1, \ldots, X_{t_n} = x_n, X_t = x) = \mathrm{P}(X_{t+\Delta} = y | X_t = x).$$

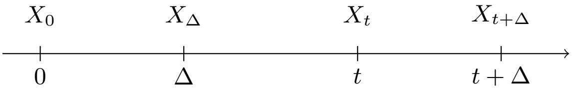
Under some regularity conditions, we then have that

$$\lim_{\delta \to 0^+} \frac{\mathrm{P}(X_{t+\delta} = y | X_t = x) - I(x, y)}{\delta} = Q_t(x, y).$$

This time-dependent matrix Q_t is called the transition rate matrix, as

$$\mathrm{P}(X_{t+\delta} = y | X_t = x) \approx I(x, y) + \delta Q_t(x, y).$$

Precise continuous-time Markov chains
Homogeneity



A CTMC is *homogeneous* or stationary if the transition rate matrix does not depend on time, that is if

$$Q_t = Q \text{ for all } t \in \mathbb{R}_{\geq 0}.$$

An alternative but equivalent condition is that

$$\mathrm{P}(X_{t+\Delta} = y | X_t = x) = \mathrm{P}(X_\Delta = y | X_0 = x).$$

Precise continuous-time Markov chains

Homogeneity

A CTMC is *homogeneous* or stationary if the transition rate matrix does not depend on time, that is if

$$Q_t = Q \text{ for all } t \in \mathbb{R}_{\geq 0}.$$

An alternative but equivalent condition is that

$$\mathrm{P}(X_{t+\Delta} = y | X_t = x) = \mathrm{P}(X_\Delta = y | X_0 = x).$$

In practice, we define a homogeneous CTMC by specifying

- its state space $\mathscr{X}$;
- its initial distribution π_0, such that $\pi_0(x) = \mathrm{P}(X_0 = x)$;
- its transition rate matrix Q.

How do we get from (π_0 and) Q to $\mathrm{E}(f(X_t)|X_0 = x)$ for some $f\colon \mathscr{X} \to \mathbb{R}$, $t \in \mathbb{R}_{\geq 0}$ and $x \in \mathscr{X}$?

How do we get from (π_0 and) Q to $\mathrm{E}(f(X_t)|X_0 = x)$ for some $f\colon \mathscr{X} \to \mathbb{R}$, $t \in \mathbb{R}_{\geq 0}$ and $x \in \mathscr{X}$?

It is well-known that

$$\mathrm{E}(f(X_t)|X_0 = x) = [T_t f](x),$$

where for all $t \in \mathbb{R}_{\geq 0}$ the matrix T_t is the solution to the *backward Kolmogorov differential equation*

$$\frac{\mathrm{d}}{\mathrm{d}t} T_t = Q T_t \text{ with initial condition } T_0 \coloneqq I.$$

How do we get from (π_0 and) Q to $\mathrm{E}(f(X_t)|X_0 = x)$ for some $f\colon \mathscr{X} \to \mathbb{R}$, $t \in \mathbb{R}_{\geq 0}$ and $x \in \mathscr{X}$?

It is well-known that

$$\mathrm{E}(f(X_t)|X_0 = x) = [T_t f](x),$$

where for all $t \in \mathbb{R}_{\geq 0}$ the matrix T_t is the solution to the *backward Kolmogorov differential equation*

$$\frac{\mathrm{d}}{\mathrm{d}t} T_t = Q T_t \text{ with initial condition } T_0 := I.$$

The solution to this differential equation is given by the matrix exponential

$$T_t = e^{tQ} := \lim_{n \to +\infty} \sum_{k=0}^{n} \frac{t^k Q^k}{k!} = \lim_{n \to +\infty} \left(I + \frac{t}{n} Q\right)^n.$$

In general, $T_t = e^{tQ}$ has to be numerically approximated!

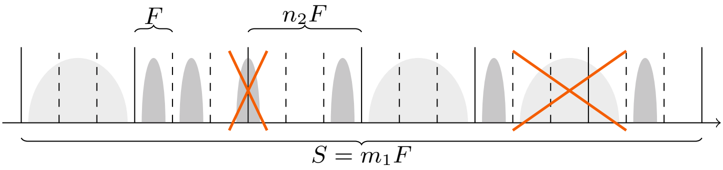
Modelling the spectrum assignment in a two-service optical grid

We consider an optical link that serves to transmit type 1 and type 2 messages.

The total available frequency spectrum S is divided into m_1 separate channels. Arriving messages are assigned to a number of contiguous channels:

- type 1 messages require 1 channel,
- type 2 messages require n_2 channels.

We combine n_2 contiguous channels into superchannels as follows



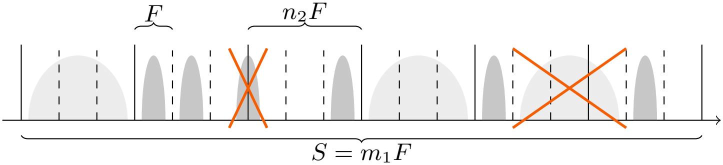
Modelling the spectrum assignment in a two-service optical grid

We consider an optical link that serves to transmit type 1 and type 2 messages.

The total available frequency spectrum S is divided into m_1 separate channels.
Arriving messages are assigned to a number of contiguous channels:

- type 1 messages require 1 channel,
- type 2 messages require n_2 channels.

We combine n_2 contiguous channels into superchannels as follows



An incoming type 1 (or type 2) message is **blocked** if it cannot be assigned to a free (super)channel.

How should we assign incoming messages in order to keep the number of blocked messages as low as possible?
We compare three assignment policies: random (R), least-filled (L) and most-filled (M).

Modelling the spectrum assignment in a two-service optical grid

Detailed state description

If we assume Poisson arrivals and exponentially distributed service times, the optical grid can be exactly modelled by a homogeneous CTMC with state space

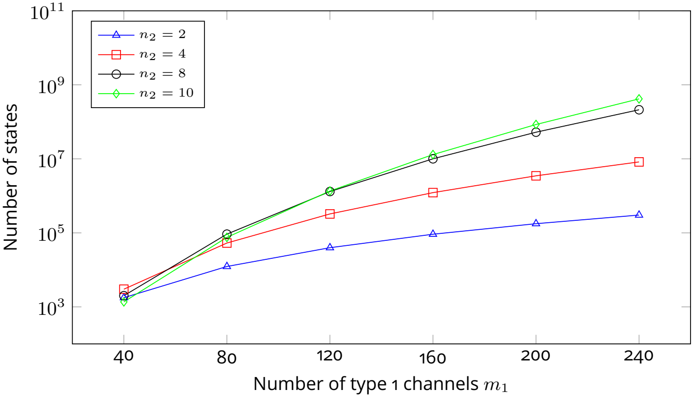
$$\mathscr{X}_{\mathrm{det}} := \Big\{ (i_0, \ldots, i_{n_2}) \in \mathbb{N}^{(n_2+1)} \colon \sum_{k=0}^{n_2} i_k \leq (m_1/n_2) \Big\},$$

where i_k counts the superchannels that are currently assigned k type 1 messages.

If we assume Poisson arrivals and exponentially distributed service times, the optical grid can be exactly modelled by a homogeneous CTMC with state space

$$\mathscr{X}_{\det} := \left\{ (i_0, \ldots, i_{n_2}) \in \mathbb{N}^{(n_2+1)} : \sum_{k=0}^{n_2} i_k \leq (m_1/n_2) \right\},$$

where i_k counts the superchannels that are currently assigned k type 1 messages.



State space explosion

Under some assumptions, many systems can be “exactly” modelled by a homogeneous CTMC. Unfortunately, the number of required states often grows exponentially with the dimensions of the system.

This makes numerically approximating e^{tQ} **infeasible** if not **impossible**.

Key to our **solution** is that many inferences only require a *reduced* or less-detailed state description.

Lumping to reduce the number of states

We **lump** (or group or aggregate) states $x, y, \ldots$ in the detailed state space $\mathscr{X}$ to "superstates" or *lumps* $\hat{x} \subset \mathscr{X}, \hat{y} \subset \mathscr{X}, \ldots$.

These lumps $\hat{x}$ form the lumped state space $\hat{\mathscr{X}}$, which is actually a partition of $\mathscr{X}$.

Lumping to reduce the number of states

We **lump** (or group or aggregate) states $x, y, \ldots$ in the detailed state space $\mathscr{X}$ to "superstates" or *lumps* $\hat{x} \subset \mathscr{X}, \hat{y} \subset \mathscr{X}, \ldots$.

These lumps $\hat{x}$ form the lumped state space $\hat{\mathscr{X}}$, which is actually a partition of $\mathscr{X}$.

Consider the lumped stochastic process $\hat{X}_t$, defined as

$$(\hat{X}_t = \hat{x}_i) \Leftrightarrow (X_t \in \hat{x}_i) \text{ for all } \hat{x} \in \hat{\mathscr{X}}.$$

Can we say something about the dynamics of this process?

Can we determine $\mathrm{E}(f(\hat{X}_t) | \hat{X}_0 = \hat{x})$*?*

We prove that—given an initial distribution π_0—the lumped process $\hat{X}_t$ is actually an inhomogeneous CTMC, with (time-dependent) transition rate matrix

$$\hat{Q}_t(\hat{x}, \hat{y}) = \frac{\sum_{x \in \hat{x}} \mathrm{P}(X_t = x) \sum_{y \in \hat{y}} Q(x, y)}{\sum_{x \in \hat{x}} \mathrm{P}(X_t = x)},$$

where

$$\mathrm{P}(X_t = x) = \sum_{y \in \mathscr{X}} \mathrm{P}(X_0 = y) \mathrm{P}(X_t = x | X_0 = y) = \sum_{y \in \mathscr{X}} \pi_0(y) [T_t \mathbb{I}_x](y).$$

We prove that—given an initial distribution π_0—the lumped process $\hat{X}_t$ is actually an inhomogeneous CTMC, with (time-dependent) transition rate matrix

$$\hat{Q}_t(\hat{x}, \hat{y}) = \frac{\sum_{x \in \hat{x}} \mathrm{P}(X_t = x) \sum_{y \in \hat{y}} Q(x, y)}{\sum_{x \in \hat{x}} \mathrm{P}(X_t = x)},$$

where

$$\mathrm{P}(X_t = x) = \sum_{y \in \mathscr{X}} \mathrm{P}(X_0 = y) \mathrm{P}(X_t = x | X_0 = y) = \sum_{y \in \mathscr{X}} \pi_0(y) [T_t \mathbb{I}_x](y).$$

As we cannot (or do not want) to determine $T_t = e^{tQ}$, we cannot determine $\hat{Q}_t$.

Current literature: only consider lumping in case the lumped chain $\hat{X}_t$ is actually homogeneous, i.e., if

$$\hat{Q}_t = \hat{Q} \text{ for all } t \in \mathbb{R}_{\geq 0}.$$

This is the case (regardless of the initial distribution π_0) if and only if for all $\hat{x}, \hat{y} \in \hat{\mathscr{X}}$,

$$\sum_{y \in \hat{y}} Q(x, y) = \sum_{y \in \hat{y}} Q(x', y) \text{ for all } x, x' \in \hat{x}.$$

In all other cases, all we can say is that for all $t \in \mathbb{R}_{\geq 0}$ and all $\hat{x}, \hat{y} \in \hat{\mathscr{X}}$,

$$\min_{x \in \hat{x}} \sum_{y \in \hat{y}} Q(x, y) \leq \hat{Q}_t(\hat{x}, \hat{y}) \leq \max_{x \in \hat{x}} \sum_{y \in \hat{y}} Q(x, y).$$

Therefore, we consider the set $\mathbb{P}_{Q, \hat{\mathscr{X}}}$ of inhomogeneous CTMCs on $\hat{\mathscr{X}}$ with initial distribution

$$\hat{\pi}_0(\hat{x}) = \sum_{x \in \hat{x}} \pi_0(x) \text{ for all } x \in \mathscr{X}$$

and time-dependent transition rate matrix $\hat{R}_t$ such that

$$\min_{x \in \hat{x}} \sum_{y \in \hat{y}} Q(x, y) \leq \hat{R}_t(\hat{x}, \hat{y}) \leq \max_{x \in \hat{x}} \sum_{y \in \hat{y}} Q(x, y) \text{ for all } \hat{x}, \hat{y} \in \hat{\mathscr{X}}.$$

Then

$$\underline{\mathrm{E}}(f(\hat{X}_{t+\Delta}) | \hat{X}_t = \hat{x}) := \min_{\mathrm{P} \in \mathbb{P}_{Q, \hat{\mathscr{X}}}} \mathrm{E}(f(\hat{X}_{t+\Delta}) | \hat{X}_t = \hat{x}) \leq \mathrm{E}(f(X_{t+\Delta}) | X_t \in \hat{x}).$$

The lumped chain naturally leads to an imprecise CTMC

The set $\mathbb{P}_{Q,\hat{\mathscr{X}}}$ actually characterises an imprecise CTMC, such that

$$\underline{\mathrm{E}}(f(\hat{X}_{t+\Delta})|\hat{X}_t = \hat{x}) = \min_{\mathrm{P}\in\mathbb{P}_{Q,\hat{\mathscr{X}}}} \mathrm{E}(f(\hat{X}_{t+\Delta})|\hat{X}_t = \hat{x})$$

can be numerically approximated pretty efficiently.

For more information, see

Thomas Krak, Jasper De Bock and Arno Siebes.
Imprecise continuous-time Markov chains.
International Journal of Approximate Reasoning, **88**:452–528, 2017.

To determine the blocking probabilities, it suffices to consider the reduced state space

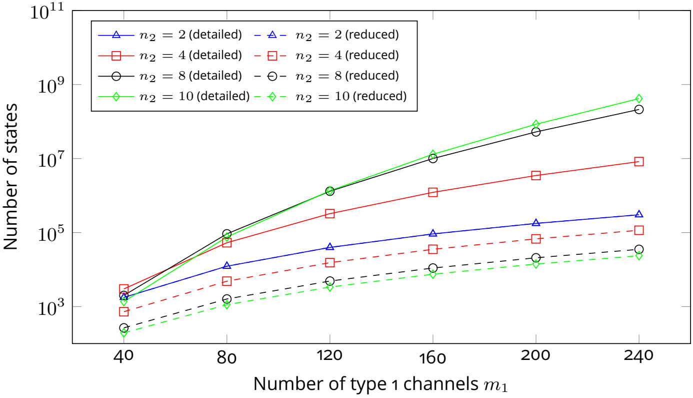
$$\mathscr{X}_{\mathrm{red}} \coloneqq \left\{(i,j,e) \in \mathbb{N}^3 \colon m_1/n_2 \leq i+j+e, i+(j+e)n_2 \leq m_1\right\},$$

where i counts the assigned type 1 messages, j counts the assigned type 2 messages and e counts the empty superchannels.

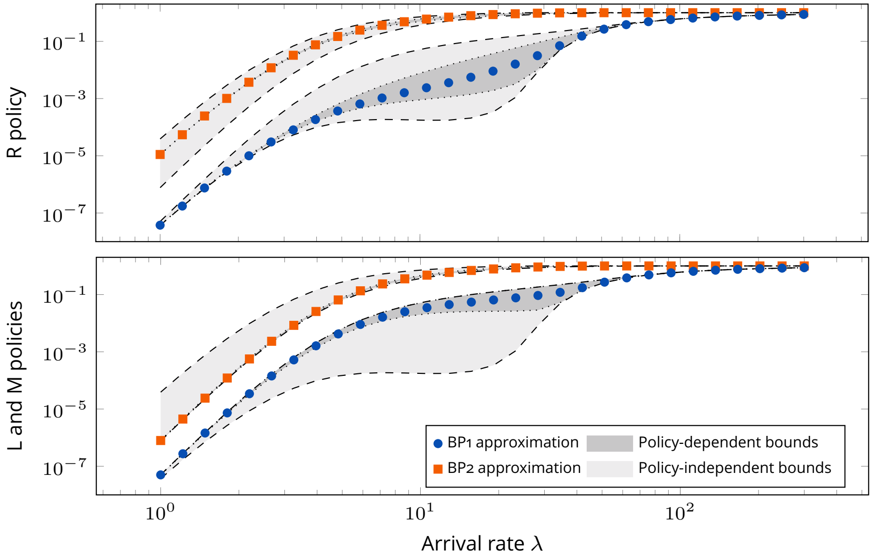
To determine the blocking probabilities, it suffices to consider the reduced state space

$$\mathscr{X}_{\mathrm{red}} \coloneqq \left\{(i,j,e) \in \mathbb{N}^3 \colon m_1/n_2 \leq i+j+e, i+(j+e)n_2 \leq m_1\right\},$$

where i counts the assigned type 1 messages, j counts the assigned type 2 messages and e counts the empty superchannels.



We compare our bounds on the blocking probabilities for a system with $m_1 = 40$, $n_2 = 4$, $\mu_1 = \mu_2 = 1$ and $\lambda_1 = \lambda_2 = \lambda$ with some precise approximation.



Some remaining questions

Want to join our *imprecise continuous-time Markov chain fanclub*?

Are there any other examples of CTMC (or DTMC) models where this approach would be useful?

Is there another field (or set of results) of imprecise probabilities that we could popularise as a computational tool instead of as a philosophically more appealing theory?