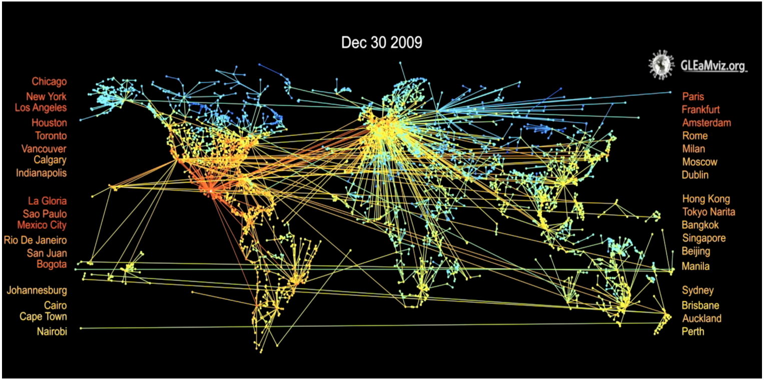
Introduction to Mean-Field

Luca Bortolussi   Nicolas Gast

Dipartimento di Matematica e Geoscienze
Università degli studi di Trieste
CNR/ISTI, Pisa
luca@dmi.units.it

SFM, Bertinoro, June 20-24, 2016

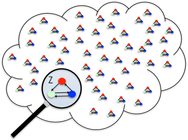
Predicting Pandemics

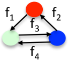


Worldwide model of H1N1 2009 influenza virus.

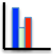
Motivational overview

- Population models are the natural way to describe large systems of interacting agents, common in systems biology, epidemiology, ecology, computer performance, CAS, …
- Such models are typically described as a stochastic process in discrete space and continuous time.





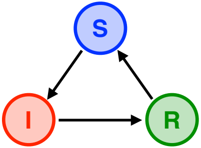
$$\frac{dx_1(t)}{dt} = F_1(x_1(t), x_2(t), x_3(t))$$
$$\frac{dx_2(t)}{dt} = F_2(x_1(t), x_2(t), x_3(t))$$
$$\frac{dx_3(t)}{dt} = F_3(x_1(t), x_2(t), x_3(t))$$

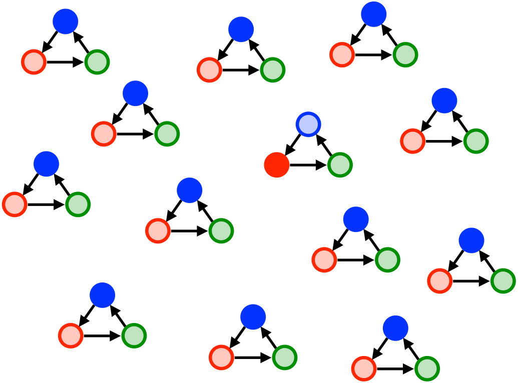


- These stochastic models are very hard to analyse, in particular for large populations.

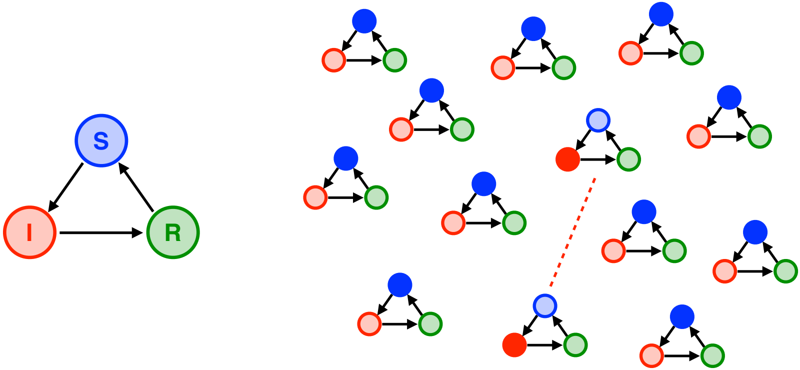
- Mean field theorems, when they can be applied, allow us to replace for large populations the stochastic process with a small system of ODE, that can be easily solved numerically.

Example: SIRS epidemic model

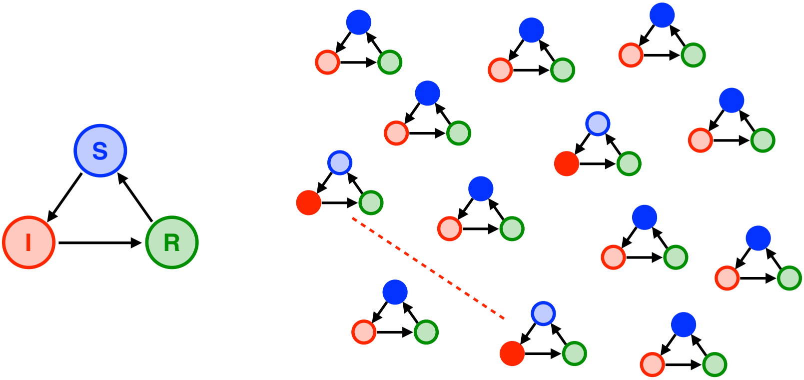




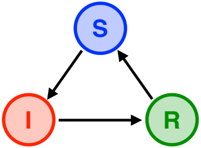
Example: SIRS epidemic model

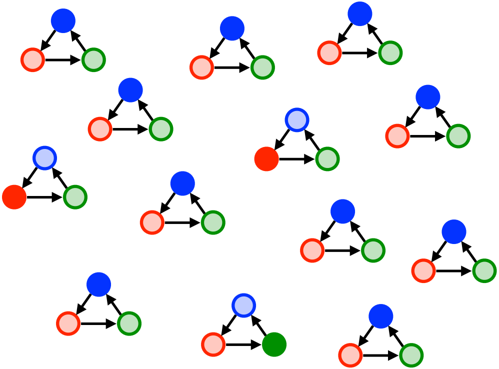


Example: SIRS epidemic model

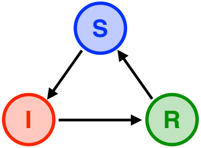


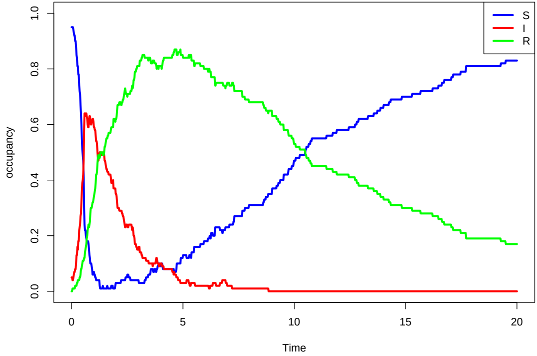
Example: SIRS epidemic model



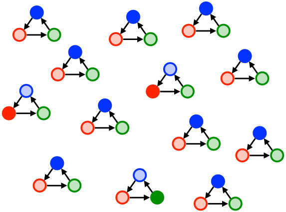


Example: SIRS epidemic model





Population Models



- Assumption: agents are individually indistinguishable and homogeneously mixed.
- State: we just need to count how many agents are in each different state.
- Dynamics: what are the interactions among agents, how many interact and how they change state.

A population CTMC model is a tuple $\mathcal{X} = (\mathbf{X}, \mathcal{D}, \mathcal{T}, \mathbf{x_0})$, where:

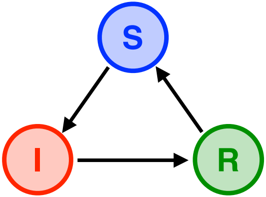
1. $\mathbf{X}$ — vector of *variables* counting how many individuals in each state.
2. $\mathcal{D}$ — (countable) state space.
3. $\mathbf{x_0} \in \mathcal{D}$ —*initial state*.
4. $\eta_i \in \mathcal{T}$ — *global transitions*. They can be visualised as chemical reaction/ rewriting rules:

$$r_1 X_1 + \ldots r_n X_n \longrightarrow s_1 X_1 + \ldots s_n X_n.$$

   Formally, they are pairs $\eta_i = (\mathbf{v}, r(\mathbf{X}))$

   4.1 $\mathbf{v} \in \mathbb{R}^n$, $\mathbf{v} = \mathbf{s} - \mathbf{r}$ — *update vector* (state changes from $\mathbf{X}$ to $\mathbf{X} + \mathbf{v}$)
   4.2 $r : \mathcal{D} \to \mathbb{R}_{\geq 0}$ — rate function.

Example: SIRS epidemics



- Three variables: X_S,X_I,X_R.
- State space: $\mathcal{D} = \{(n_1, n_2, n_3) \mid n_1 + n_2 + n_3 = N\} \subset \{0, \ldots, N\}^3$.
- Transitions:
  - $S + I \longrightarrow I + I$
    $\eta_{inf} = ((-1, 1, 0), sk_I \frac{X_I}{N} X_S)$
  - $I \longrightarrow R$
    $\eta_{rec} = ((0, -1, 1), k_R X_I)$
  - $R \longrightarrow S$
    $\eta_{susc} = ((1, 0, -1), k_S X_R)$

Exponential Distribution

Definition

A random variable $T : (\Omega, \mathcal{S}) \to [0, \infty]$ is $Exp(\lambda)$ iff

- Cdf is $\mathbb{P}(T < t) = 1 - e^{-\lambda t}$
- Density is $f_T(t) = \lambda e^{-\lambda t}$, $t \geq 0$.

The expected value of T is $\mathbb{E}(T) = \int_0^\infty \mathbb{P}(T > t)dt = \frac{1}{\lambda}$.

Memoryless Property

$T \sim Exp(\lambda)$ if and only if the following memoryless property holds:

$$\mathbb{P}(T > s + t | T > s) = \mathbb{P}(T > t) \text{ for all } s, t \geq 0.$$

Instantaneous firing probability

An exponential distribution with rate λ models the firing time of an event who has probability of firing between time t and $t + dt$ equal to λdt.

CTMC: definition

quanticol
www.quanticol.eu

S-valued Continuous Time Markov Chain

- Let S be finite or countable.
- A CTMC on a state space S is a labelled graph, where labels are the rates of exponential distributions.
- In each state, there is a race condition between the different exiting edges: the fastest is traversed.
- The CTMC has the memoryless property: the future depends only on the current state.

Formally

A Continuous Time Markov Chain is a right-continuous continuous-time random process (with cadlag sampling paths) satisfying the memoryless condition: for each n, t_i and s_i:

$$\mathbb{P}(X_{t_n} = s_n \mid X_{t_0} = s_0, \ldots, X_{t_{n-1}} = s_{n-1}) = \mathbb{P}(X_{t_n} = s_n \mid X_{t_{n-1}} = s_{n-1}).$$

CTMC: infinitesimal generator

quanticol
www.quanticol.eu

Formally

A Continuous Time Markov Chain is a right-continuous continuous-time random process (with cadlag sampling paths) satisfying the memoryless condition: for each n, t_i and s_i:

$$\mathbb{P}(X_{t_n} = s_n \mid X_{t_0} = s_0, \ldots, X_{t_{n-1}} = s_{n-1}) = \mathbb{P}(X_{t_n} = s_n \mid X_{t_{n-1}} = s_{n-1}).$$

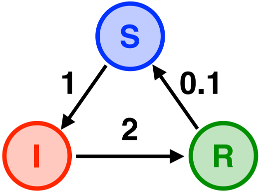
Q-matrix

A Q-matrix is the $|S| \times |S|$ matrix such that:

1. $q_{ij} \geq 0$, $i \neq j$ is the rate of the exponential distribution giving the time needed to go from state s_i to state s_j
2. $q_{ii} = -\sum_{j \neq i} q_{ij}$ is the opposite of the exit rate from state i.

Therefore, each row of the Q-matrix sums up to zero.

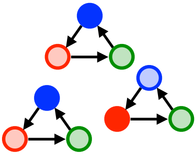
A simple example: single agent infection

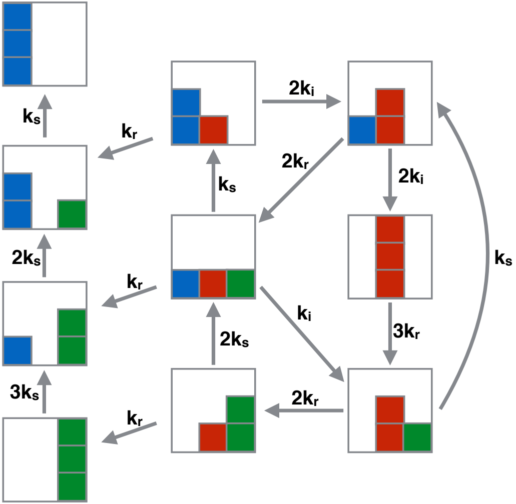


$$S = \{S, I, R\}$$

$$Q = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -2 & 2 \\ 0.1 & 0 & -0.1 \end{pmatrix}$$

CTMC for population models: the SIRS example





Master Equation

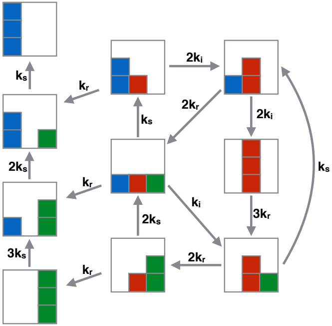
The equation for the time evolution of the probability mass for CTMC is known as Kolmogorov equation. In the context of Population Processes is often know as master equation.

There is one equation per state $\mathbf{x} \in \mathcal{D}$, for the probability mass $P(\mathbf{x}, t)$, which considers the inflow and outflow of probability at time t.

$$\frac{dP(\mathbf{x}, t)}{dt} = \sum_{\eta \in \mathcal{T}} r_\eta(\mathbf{x} - \mathbf{v}_\eta) P(\mathbf{x} - \mathbf{v}_\eta, t) - \sum_{\eta \in \mathcal{T}} r_\eta(\mathbf{x}) P(\mathbf{x}, t)$$

These differential equations, for finite state spaces, can be solved by numerical integration or by using specialised methods for CTMC (uniformization). Finite state projections can be used for infinite state spaces. The cost is polynomial in the size of the state space.

Example: SIRS model



$$dP([1,1,1],t)/dt = 2k_r P([1,2,0],t) + 2k_s P([0,1,2],t) - k_i P([1,1,1],t) - k_r P([1,1,1],t) - k_s P([1,1,1],t)$$

$$dP([3,0,0],t)/dt = k_s P([2,0,1],t)$$

$$dP([0,3,0],t)/dt = 2k_i P([1,2,0],t) - 3k_r P([0,3,0],t)$$

An alternative to solve the master equation is to generate sample trajectories of the CTMC, and then extract statistical information from them.

The most famous simulation algorithm is due to Doob-Gillespe. It is based on the fact that a CTMC can be factorized in two independent processes.

- The time $T_{\mathbf{x}}$ spent in a state (holding time) $\mathbf{x}$ is exponentially distributed with rate $r_0(\mathbf{x}) = \sum_\eta r_\eta(\mathbf{x})$ (exit rate).
- The probability of taking transition η (jump chain) is independent of T and is equal to $r_\eta(\mathbf{x})/r_0(\mathbf{x})$.

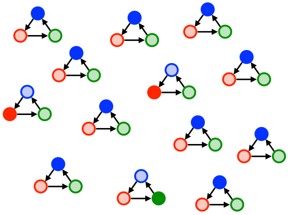
The Doob-Gillespie algorithm samples the time spent in a state and the next state according to the previous characterization. Let $\mathbf{x}$ be the current state (initially $\mathbf{x_0}$) and t the current time (initially, $t = 0$).

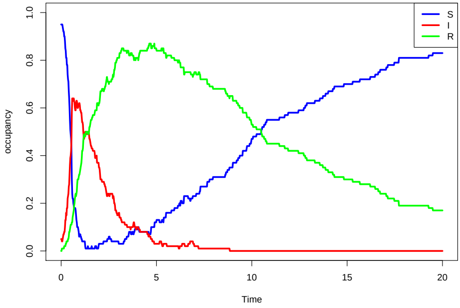
While $t < t_{final}$

1. Sample $dt \sim T_{\mathbf{x}}$ (using $dt = -\log U / r_0(\mathbf{x})$, U uniform in $[0, 1]$) and update time to $t + dt$
2. Choose a transition η with probability $r_\eta(\mathbf{x}) / r_0(\mathbf{x})$ and update the state to $\mathbf{x} + \mathbf{v}_\eta$.

Complexity is proportional on the number of reactions and on the steps to be done till the final time is reached, which on average are bounded by $t_{final} \cdot \max_{\mathbf{x}} r_0(\mathbf{x})$.

Example: SIRS model





System size

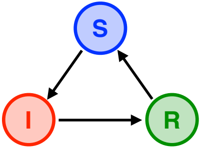
A crucial notion in population models is that of system size:

- typically the total population (e.g. for the SIRS model)
- a physical quantity (the volume in biochemical systems)
- a measure of intensity (the arrival rate in queueing networks)

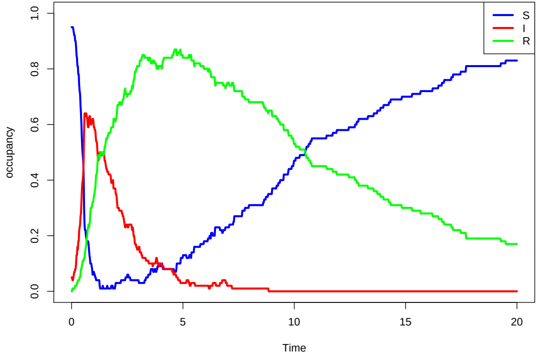
Why it is important?

The size of the state space grows polynomially (or exponentially) with the system size. For moderate system sizes, numerical solution of the master equation is unfeasible. Simulation, instead, typically has a complexity growing linearly with the size.

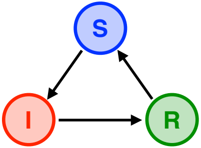
Example: SIRS epidemic model



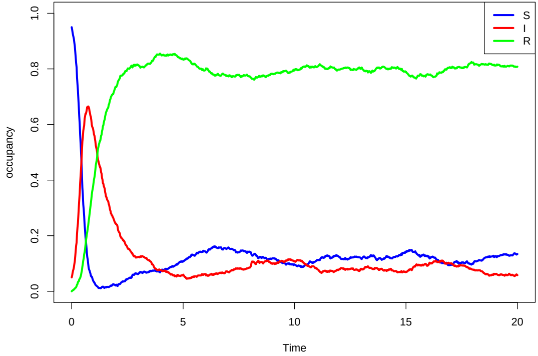
$N = 100$



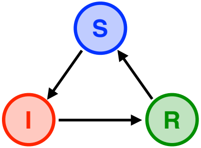
Example: SIRS epidemic model



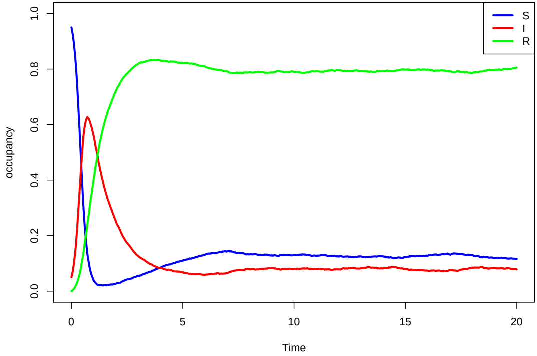
$N = 1000$



Example: SIRS epidemic model



$N = 10000$



Mean-Field (Fluid) Approximation

quanticol
www.quanticol.eu

Basics

- It applies to CTMC models of population dynamics with large population size N (studies the limit as $N \to \infty$)

Basics

- It applies to CTMC models of population dynamics with large population size N (studies the limit as $N \to \infty$)
- It works on scaled variables, to treat uniformly different population levels.

Mean-Field (Fluid) Approximation

Basics

- It applies to CTMC models of population dynamics with large population size N (studies the limit as $N \to \infty$)
- It works on scaled variables, to treat uniformly different population levels.
- Requires proper scaling and regularity assumptions on rates.

Basics

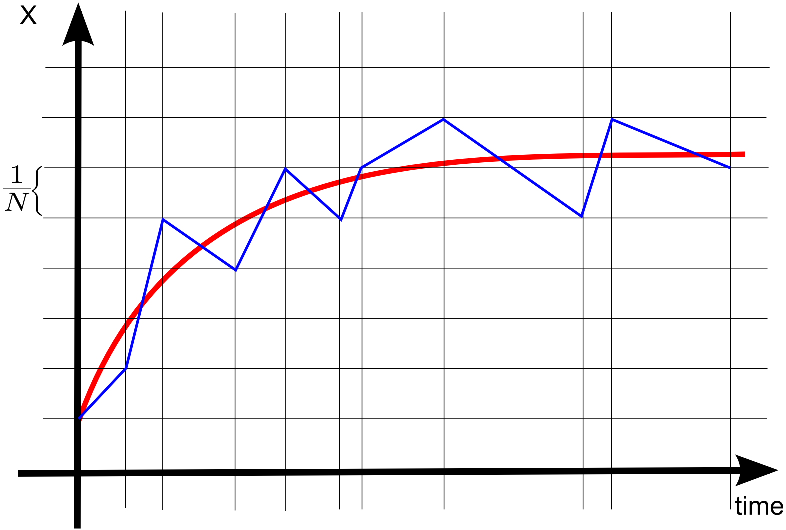
- It applies to CTMC models of population dynamics with large population size N (studies the limit as $N \to \infty$)
- It works on scaled variables, to treat uniformly different population levels.
- Requires proper scaling and regularity assumptions on rates.
- The method works by constructing an ODE from the sequence of population dependent CTMC.

Mean-Field (Fluid) Approximation

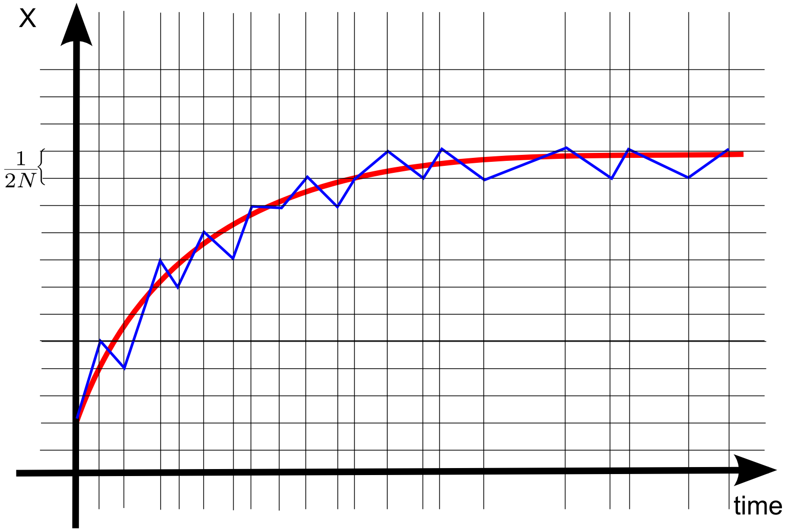
Basics

- It applies to CTMC models of population dynamics with large population size N (studies the limit as $N \to \infty$)
- It works on scaled variables, to treat uniformly different population levels.
- Requires proper scaling and regularity assumptions on rates.
- The method works by constructing an ODE from the sequence of population dependent CTMC.
- It can be proved that, in any finite time horizon, the trajectories of the CTMC become indistinguishable from the solution of the ODE.

An intuition



An intuition



Scaling Conditions

Basics

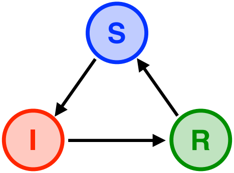
- We have a sequence $\mathcal{X}^{(N)}$ of models, for increasing system size (e.g. total population N).
- We normalize such models in order to bring them to the same scale (divide variables by size N).
- $\mathbf{X}^{(N)}(t)$ is the Markov process (in continuous time) defined by $\mathcal{X}^{(N)}$.

Normalization

The normalized model $\hat{\mathcal{X}}^{(N)} = (\hat{\mathbf{X}}, \hat{\mathcal{D}}^{(N)}, \hat{\mathcal{T}}^{(N)}, \hat{\mathbf{X}}_0^{(N)})$ associated with $\mathcal{X}^{(N)} = (\mathbf{X}, \mathcal{D}^{(N)}, \mathcal{T}^{(N)}, \mathbf{X}_0^{(N)})$ is defined by:

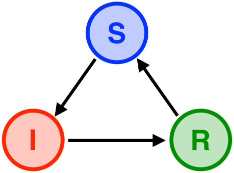
- Variables: $\hat{\mathbf{X}} = \frac{\mathbf{X}}{N}$
- Domain: $\hat{\mathcal{D}}^{(N)} = \{N^{-1}\mathbf{x} \mid \mathbf{x} \in \mathcal{D}\}$.
- Initial conditions: $\hat{\mathbf{X}}_n^{(N)} = \frac{\mathbf{X}_0^{(N)}}{N}$
- Normalized transition $\hat{\tau} = (\hat{\mathbf{v}}_\tau^N, \hat{r}_\tau^{(N)}(\hat{\mathbf{X}}))$ associated with $\tau \in \mathcal{T}^{(N)}$:
  - Update: $\hat{\mathbf{v}}_\tau^N = \mathbf{v}_\tau / N$;
  - Rates: $r_\tau^{(N)}(\mathbf{X}) = N \cdot f_\tau^{(N)}\left(\frac{\mathbf{X}}{N}\right) = \hat{r}_\tau^{(N)}\left(\frac{X}{N}\right)$

Example: SIRS epidemics



- $r_{rec}^{(N)}(\mathbf{X}) = k_R X_I = N k_R \frac{X_I}{N} = N k_R \hat{X}_I$
  $\hat{r}_{rec}^{(N)}(\hat{\mathbf{X}}) = N k_R \hat{X}_I$, $f_{rec}(\hat{\mathbf{X}}) = k_R \hat{X}_I$

Example: SIRS epidemics



- $r^{(N)}_{rec}(\mathbf{X}) = k_R X_I = N k_R \frac{X_I}{N} = N k_R \hat{X}_I$
  $\hat{r}^{(N)}_{rec}(\hat{\mathbf{X}}) = N k_R \hat{X}_I,\ f_{rec}(\hat{\mathbf{X}}) = k_R \hat{X}_I$

- $r^{(N)}_{inf}(\mathbf{X}) = \frac{k_I}{N} X_S X_I = N k_I \frac{X_S}{N} \frac{X_I}{N} = N k_I \hat{X}_S \hat{X}_I$
  $\hat{r}^{(N)}_{inf}(\hat{\mathbf{X}}) = N k_I \hat{X}_S \hat{X}_I,\ f_{inf}(\hat{\mathbf{X}}) = k_I \hat{X}_S \hat{X}_I$

Scaling assumptions: state space

- Consider the normalised state space $\hat{\mathcal{D}}^{(N)}$ of $\hat{\mathbf{X}}^{(N)}(t)$.

Scaling assumptions: state space

- Consider the normalised state space $\hat{\mathcal{D}}^{(N)}$ of $\hat{\mathbf{X}}^{(N)}(t)$.
- We need to find a set $E \subset \mathbb{R}^n$ (open or compact) which contains $\hat{\mathcal{D}}^{(N)}$ for each N. This will be the set in which the fluid limit will live.

Scaling assumptions: state space

- Consider the normalised state space $\hat{\mathcal{D}}^{(N)}$ of $\hat{\mathbf{X}}^{(N)}(t)$.
- We need to find a set $E \subset \mathbb{R}^n$ (open or compact) which contains $\hat{\mathcal{D}}^{(N)}$ for each N. This will be the set in which the fluid limit will live.

Example: SIRS epidemics

In this case, the normalised variables take values in a discrete grid between 0 and 1:

$$\hat{\mathcal{D}}_i^{(N)} = \{\frac{j}{N} \mid j = 1, \ldots, N\}.$$

Hence, we can take E to be the unit cube $[0,1]^3$.

However, the total population is conserved, so we can restrict to the unit simplex $E = \{\mathbf{x} \in [0,1]^3 \mid \sum_i x_i = 1\}$.

Scaling assumptions

$f_\tau^{(N)}$ is required to converge uniformly to a locally Lipschitz continuous and locally bounded function f_τ:

$$\sup_{\mathbf{x}\in E} \|f_\tau^{(N)}(\mathbf{x}) - f_\tau(\mathbf{x})\| \to 0.$$

If $f_\tau^{(N)} = f_\tau$ does not depend on N, the rate satisfies the density dependence condition.

f locally Lipschitz iff $\forall \mathbf{x}, \exists B(\mathbf{x}), L > 0, \forall \mathbf{y} \in B(\mathbf{x})\ \|f(\mathbf{x}) - f(\mathbf{y})\| \le L\|\mathbf{x} - \mathbf{y}\|$
f locally bounded iff $\forall \mathbf{x}, \exists B(\mathbf{x}), M > 0,\ \|f(\mathbf{x})\| \le M\|\mathbf{x} - \mathbf{y}\|$

The following theorem works also under less restrictive assumptions (e.g. random increments with bounded variance and average).

Drift and Limit Vector Field

Drift

The drift or mean increment at level N is

$$F^{(N)}(\mathbf{x}) = \sum_{\tau \in \mathcal{T}} \mathbf{v}_\tau f_\tau^{(N)}(\mathbf{x})$$

By the scaling assumptions, $F^{(N)}$ converges uniformly to F, the limit vector field:

$$F(\mathbf{x}) = \sum_{\tau \in \mathcal{T}} \mathbf{v}_\tau f_\tau(\mathbf{x}).$$

Fluid ODE

The fluid ODE is

$$\frac{d\mathbf{x}(t)}{dt} = F(\mathbf{x}(t))$$

Deterministic approximation theorem

- $\hat{\mathbf{X}}^{(N)}(t)$: sequence of Markov processes that satisfy the conditions above.
- $\exists \mathbf{x_0} \in S$ such that $\hat{\mathbf{X}}^{(N)}(0) \to \mathbf{x_0}$ in probability (or almost surely)
- $\mathbf{x}(t)$: solution of $\dot{\mathbf{x}} = F(\mathbf{x})$, $\mathbf{x}(0) = \mathbf{x_0}$, living in E for all $t \geq 0$.

Theorem (Kurtz)

For any finite time horizon $T < \infty$, it holds that:

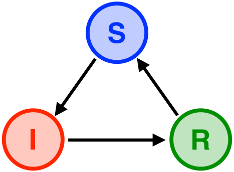
$$\sup_{0 \leq t \leq T} ||\hat{\mathbf{X}}^{(N)}(t) - \mathbf{x}(t)|| \to 0 \textit{ in probability},$$

meaning, for each $\delta > 0$, that

$$\lim_{N \to \infty} \mathbb{P}\left\{\sup_{0 \leq t \leq T} ||\hat{\mathbf{X}}^{(N)}(t) - \mathbf{x}(t)|| > \delta\right\} = 0.$$

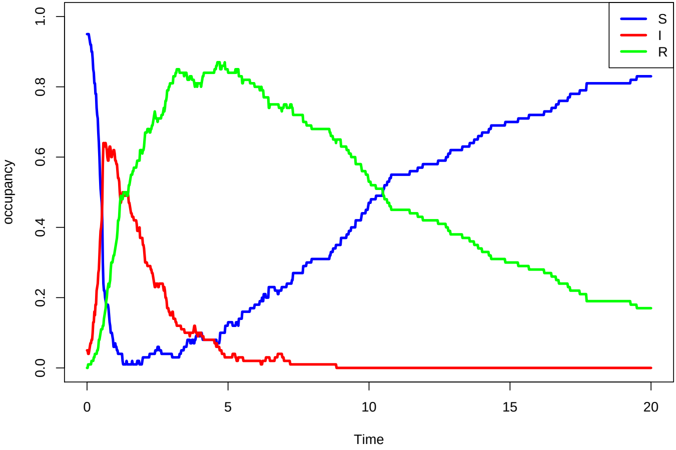
Epidemics example continued

The CTMC $\mathbf{X}^{(N)}(t)$ of the epidemics model satisfies all the hypothesis of fluid limit theorem, so it converges in probability to the solution of the following set of ODEs:



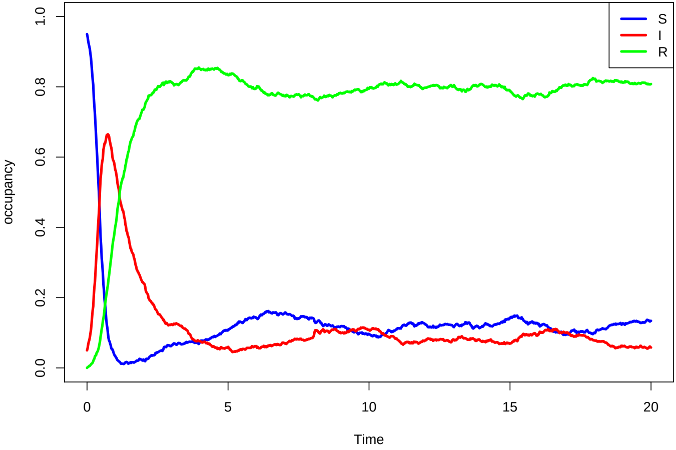
$$\begin{cases} \frac{dx_S}{dt} = k_S x_R - k_I x_I x_S \\ \frac{dx_I}{dt} = k_I x_I x_S - k_R x_I \\ \frac{dx_R}{dt} = k_R x_I - k_S x_R \end{cases}$$

Epidemics example continued



CTMC $N = 100$

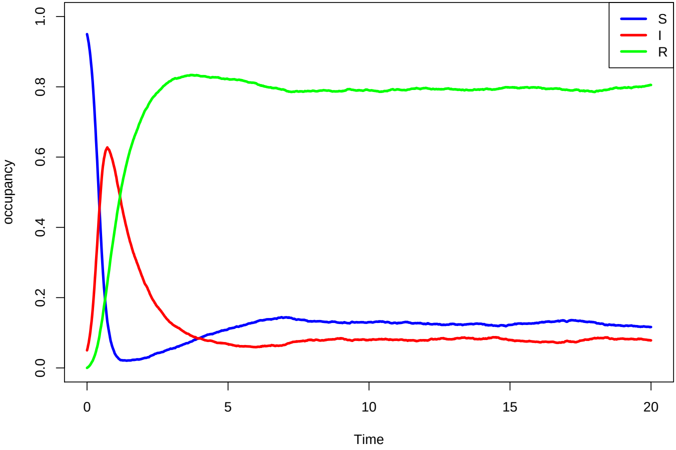
Epidemics example continued



CTMC $N = 1000$

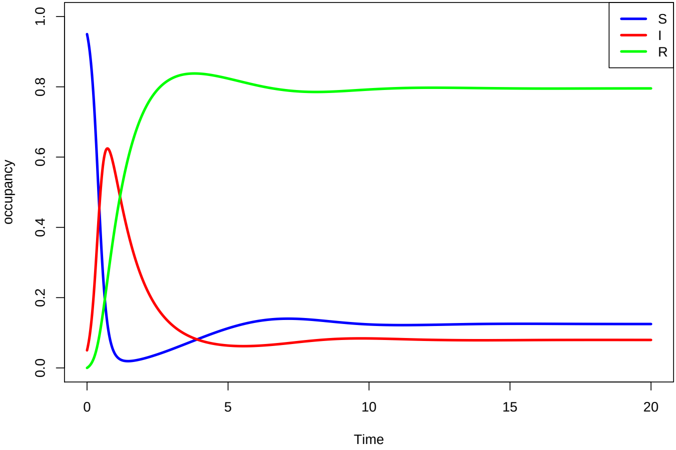
Epidemics example continued

quanticol
www.quanticol.eu



CTMC $N = 10000$

Epidemics example continued



Limit ODE

Proof of Kurtz Theorem: intuition

General idea: CTMC as a perturbed dynamical system

$$\hat{\mathbf{X}}^{(N)}(t) = \hat{\mathbf{X}}^{(N)}(0) + \int_0^t F(\hat{\mathbf{X}}^{(N)}(s))ds + M^{(N)}(t),$$

$$M^{(N)}(t) := \hat{\mathbf{X}}^{(N)}(t) - \hat{\mathbf{X}}^{(N)}(0) - \int_0^t F(\hat{\mathbf{X}}^{(N)}(s))ds$$

- Compensator: $\hat{\mathbf{X}}^{(N)}(0) + \int_0^t F(\hat{\mathbf{X}}^{(N)}(s))ds$
- ODE solution: $\mathbf{x}(t) = \mathbf{x}(0) + \int_0^t F(\mathbf{x}(s))ds$
- Noise term $M^{(N)}(t)$: $\sup_{t\leq T} \|M^{(N)}(t)\|$ converges to zero as $1/\sqrt{N}$ in probability. Staten otherwise, the magnitude of fluctuations of the non-normalised population model are of order $\sqrt{N}$.

Mean-field for Process Algebra

- The results on mean field approximation are independent of the way we specify models, provided the theorem conditions are satisfied. We can apply it to Stochastic Process Algebra models.
- SPA usually generate a CTMC via a Structural Operational Semantics (SOS). But one can define a SOS that generates the mean-field equations.
  In some cases, SPA properties automatically guarantee that the conditions of the mean-field theorem are satisfied.
- Mean-Field semantics for SPA exist for PEPA, Bio-PEPA, sCCP, stochastic CCS, . . .

quanticol
www.quanticol.eu

A simple CCS-like SPA

$$A{:=}!a.C \mid ?a.C \mid a.C \mid A{+}A; \quad C{:=}A; \quad P{:=}A \mid P \parallel P; \quad a \in \mathcal{A}, k(a) \in \mathbb{R}_{\geq 0}$$

We have two types of interaction:

- handshake synchronisation $(!a, ?a)$, with global rate proportional to the number of agent's pairs that can perform it.
- spontaneous actions (a), with global rate proportional to the number of agents that can perform it.

SIRS example

$$S{:=}?inf.I; \quad I{:=}!inf.I + rec.R; \quad R{:=}loss.S;$$

$$SIRS{:=}\underbrace{S \parallel \ldots \parallel S}_{n_S} \parallel \underbrace{I \parallel \ldots \parallel I}_{n_I}$$

Mean-field for Process Algebra

quanticol
www.quanticol.eu

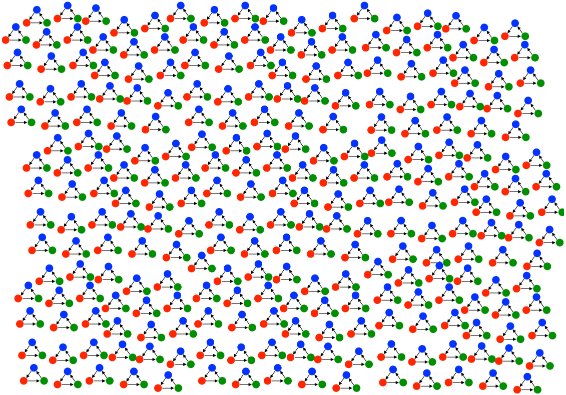
How to build the mean-field ODE via SOS?

1. Store the initial number of agents for each type in a counting vector.
2. Build the reduced system: put one single copy of each agent type in parallel.
3. Apply a set of SOS rules labeled by an update vector and a rate function.
4. collect the set of all possible derivations: these will be the transitions of a PCTMC.

Examples of SOS rules (whiteboard)

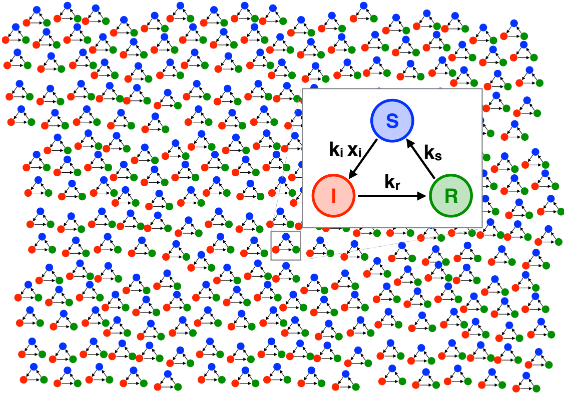
- $!a.C \xrightarrow{!a,e_C,k(a)} C$, $?a.C \xrightarrow{?a,e_C,1} C$, $a.C \xrightarrow{a,e_C,k(a)} C$.
- $P_1 \xrightarrow{!a,v_1,f_1} P_1'$, $P_2 \xrightarrow{?a,v_2,f_2} P_2' \Rightarrow P_1 \parallel P_2 \xrightarrow{a,v_1+v_2,f_1 \cdot f_2} P_1' \parallel P_2'$

Fast Simulation



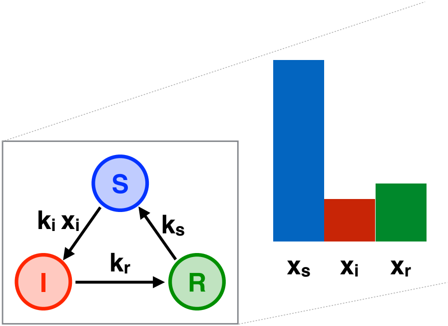
- Consider a large population

Fast Simulation



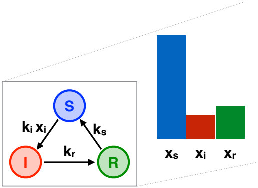
- Consider a large population
- Focus on an individual agent
- We can model it as a CTMC conditional on the global state

Fast Simulation



- Consider a large population
- Focus on an individual agent
- We can model it as a CTMC conditional on the global state
- Fast simulation: replace the PCTMC with its mean-field

Fast Simulation



The formal treatment of fast simulation requires:

- Define the model of an individual agent conditional on the system state
- Prove a convergence result, when the system is replaced by its mean-field approximation.

Single Agent Asymptotic Behaviour

- Fix a single individual in a population model $\mathcal{X}^{(N)}$ and let $Z^{(N)}$ be the single-agent stochastic process with state space S (not necessarily Markov).
- Let $Q^{(N)}(\mathbf{x})$ be defined by

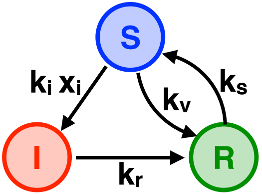
$$\mathbb{P}\{Y_h^{(N)}(t+dt) = j \mid Y_h^{(N)}(t) = i, \hat{\mathbf{X}}^{(N)}(t) = \mathbf{x}\} = q_{i,j}^{(N)}(\mathbf{x})dt,$$

  with $Q^{(N)}(\mathbf{x}) \to Q(\mathbf{x})$.
- Let $z(t)$ be the time inhomogeneous-CTMC on S with infinitesimal generator $Q(t) = Q(\mathbf{x}(t))$, $\mathbf{x}(t)$ fluid limit.

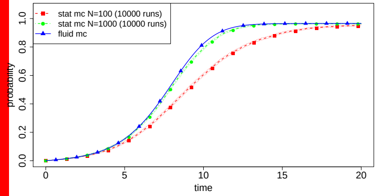
Theorem (Fast simulation theorem)

For any $T < \infty$, $\mathbb{P}\{Z^{(N)}(t) \neq z(t), t \leq T\} \to 0.$

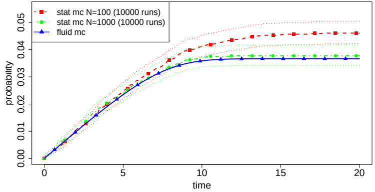
Fluid Model Checking



- Goal: check properties of an individual agent.
- Idea: model check the fast simulation model.
- Challenge: the model is a time-dependent CTMC.
- Gain: speedup of few orders of magnitude.

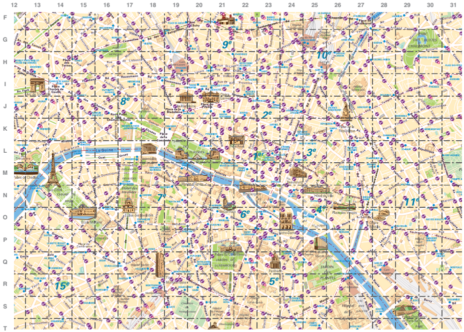


$\mathbf{X}|_{[0,T]}\texttt{infected}$



$\neg\texttt{infected}\,\mathbf{U}_{[0,T]}\texttt{vaccinated}$

Bike Sharing System



Vélib' stations in the centre of Paris

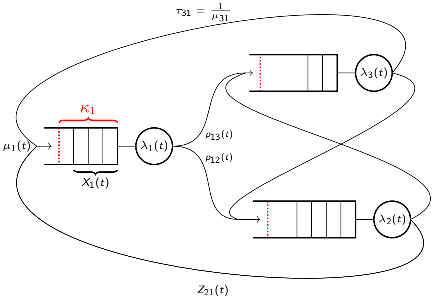
(a) Empty station

(b) Full station

- Each station has a given number of parking slots.
- Users enter the system by picking up a bike at a station, if any, and making a trip to another station, where they drop the bike on an available parking spot, if any.

The queueing model

A BSS network with 3 stations

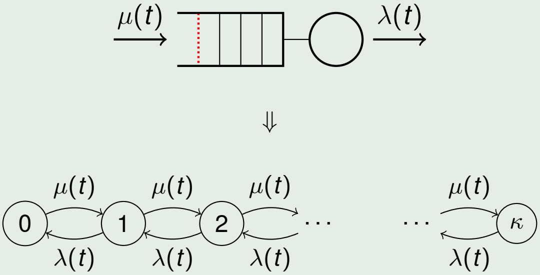


- Assume a fully symmetric situation: each station has k slots, arrival rates are the same, routing is uniform.
- Each station can be seen as an agent with internal states $\{0, \ldots, k\}$.
- We can build a population model with counting variables $X_0, \ldots, X_k$.

- Transitions: arrival of a customer in a station with i bikes $(X_i - 1, X_{i-1} + 1)$, with rate $\lambda(t)X_i(t)$, and the return of a bike from station i to station with j bikes $(X_j - 1, X_{j+1} + 1)$ with rate $\mu_{ij}(t)X_j(t)$.
- System size: number of stations N. We can apply the mean field limit.

Mean field analysis

Fast simulation → treat stations independently



Beyond homogeneity

This approach works if all stations are the same, which is not realistic.

It can be adapted to the case of heterogeneous stations (next part of the talk)!

Summary

- Context: stochastic models of large populations of interacting agents (e.g. epidemics, bike sharing, . . .).
- Problem: hard to analyse computationally.
- Main idea: theorems showing that under mild regularity conditions, the behaviour of these models for large populations is essentially captured by a small set of ODEs.
- Pros: fast computational methods to analyse global behavior, and to check properties of individuals.
- Cons: limitations due to the conditions to be satisfied (large populations, continuous rates, scaling conditions, homogeneity). Relaxations are possible.

Take home message

If you need to analyse a large population model, check if you can apply mean-field. If so, use it and save a lot of time and energy.

Bibliography

- L. Bortolussi, J. Hillston, D. Latella, and M. Massink, “Continuous approximation of collective system behaviour: A tutorial”, Performance Evaluation, vol. 70, no. 5, pp. 317-349, May 2013.
- R. W. R. Darling and J. R. Norris, “Differential equation approximations for Markov chains.., Probability Surveys, vol. 5, pp. 37-79, 2008.
- L. Bortolussi and J. Hillston, “Model checking single agent behaviours by fluid approximation”, Information and Computation, vol. 242, pp. 183-226, Jun. 2015.
- N. Gast, G. Massonnet, D. Reijsbergen, and M. Tribastone, “Probabilistic forecasts of bike-sharing systems for journey planning”, 24th ACM International Conference on Information and Knowledge Management (CIKM 2015), 2015.