Approximation and statistical inference in stochastic models of reaction (and interaction) networks, T4

Bibbona
Gasparini, Erhardt, Como, Vaccarino, . . .

Torino, 5-3-2018

Reaction networks

Different models in biology, chemistry, physics, epidemiology, neuroscience, ecology and social sciences can be presented as a set of reactions (interactions) among chemical species, populations, computers...

This reactions are normally organized as a graph like the following

$$A \underset{\lambda_2}{\overset{\lambda_1}{\leftrightarrows}} \emptyset \underset{\lambda_4}{\overset{\lambda_3}{\leftrightarrows}} B$$

$$2A \xleftarrow{\lambda_5} A + B \xrightarrow{\lambda_6} 2B$$

- the *chemical species*
- the *complexes*, linear combinations of the species (*vertices*)
- the *reactions*, rules how to convert a complex into another (*edges*)

Different (stochastic) modelling paradigms

$$\emptyset \underset{\lambda_2}{\overset{\lambda_1}{\leftrightarrows}} B \underset{\lambda_4}{\overset{\lambda_3}{\leftrightarrows}} O$$

reaction networks allow for automatic translation into

- Markov chain models. Works at every scale. It becomes numerically unmanageable for large systems (many molecules, individuals, agents...)
- Deterministic approximations. A system of ODEs. It is a large scale limit, a good approximation for systems with many molecules, individuals, agents...
- Diffusion approximations. Keeps stochasticity. It is a good approximation for systems with intermediate size.

Some history: deterministic setting (N_A is large)

algebraic properties of the reaction graph

$\Downarrow$

dynamical properties of the system (equilibria, persistence, multistability)

Deficiency zero theorem, Feinberg '79

Deficiency $\delta = c - l - s = 0$ + weak reversibility

$\Downarrow$

for all choices of the constants, there is one equilibrium concentration in each positive stoichiometric compatibility class and it is locally asymptotically stable.

Global stability is an open problem, the Global Attractor Conjecture. Craciun claimed a proof in 2015. Necessary checks are still ongoing.

Some recent history: stochastic setting

Sometimes the theorems which hold in the deterministic setting, may have an analog in the stochastic context. Results on stochastic modeling are very recent

- D.F. Anderson, G. Craciun, and T.G. Kurtz, Product-form stationary distributions for deficiency zero chemical reaction networks, *Bulletin of Mathematical Biology*, 2010.
- D. Cappelletti, C. Wiuf, Product-form Poisson-like distributions and complex balanced reaction systems. *SIAM Journal on Applied Mathemathics,* 2016
- D.F. Anderson, D. Cappelletti, M. Koyama, and T.G. Kurtz, Non-explosivity of stochastically modeled reaction networks that are complex balanced, Arxiv, 2017.
- A. Agazzi, A. Dembo, and J.P. Eckmann, Large deviations theory for Markov jump models of chemical reaction networks, *Ann. Appl. Prob.* (2018).

Applications and relations with other disciplines

- Non equilibrium thermodinamics ▸ Link
- Systems Biology, gene expression . . . ▸ Link
- Physical realization of given reaction networks with DNA strands ▸ Link
- Engeneering or controlling reaction networks to obtain specific properties ▸ Link
- Computing with reaction networks ▸ Link

Two keywords for our project

- Approximation
- Statistical inference

Part one ends here

Part two

Classical approximations: what can go wrong

Classical approximation theorems are formulated on a finite time horizon. Asymptotically things are less clear, no general results:

- Sometimes deterministic (MC) and stochastic models are in agreement, e.g. under complex balance.
- Sometimes (e.g under Absolute Concentration Robustness) deterministic (MC) and stochastic models have different asymptotic behaviors.

E.g. Back to **Lotka-Volterra** (stochastic)
two states are special:

$$(p = \infty, P = 0) \qquad (p = 0, P = 0)$$

asymptotically, one of these two states is reached with probability 1.

On the contrary, the deterministic system will keep on oscillating forever

What is the asymptotic behavior of the diffusion approx?

- When ODE and MC are in agreement is the diffusion in agreement as well?
- When ODE and MC are in disagreement, is the diffusion approximation in agreement with the MC?
- When there is an absorbing state according to both the MC and the diffusion approximation. Are the distributions of the times to absorption in agreement? If they exist, are the pseudo stationary distributions in agreement?

Degeneracy of the diffusion at the boundary

Since concentrations cannot be negative, the plane when a component is zero is always a *boundary* of the state space. In some examples the MC can enter the boundary and then jumps back to the interior. E.g.

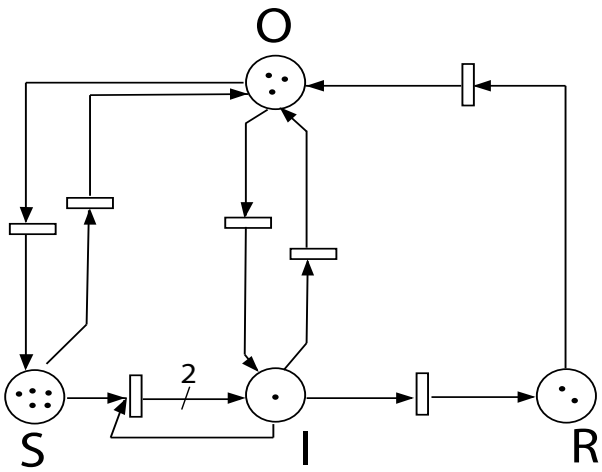
$$A \underset{\lambda}{\overset{\lambda}{\leftrightarrows}} \emptyset \underset{\lambda}{\overset{\lambda}{\leftrightarrows}} B$$

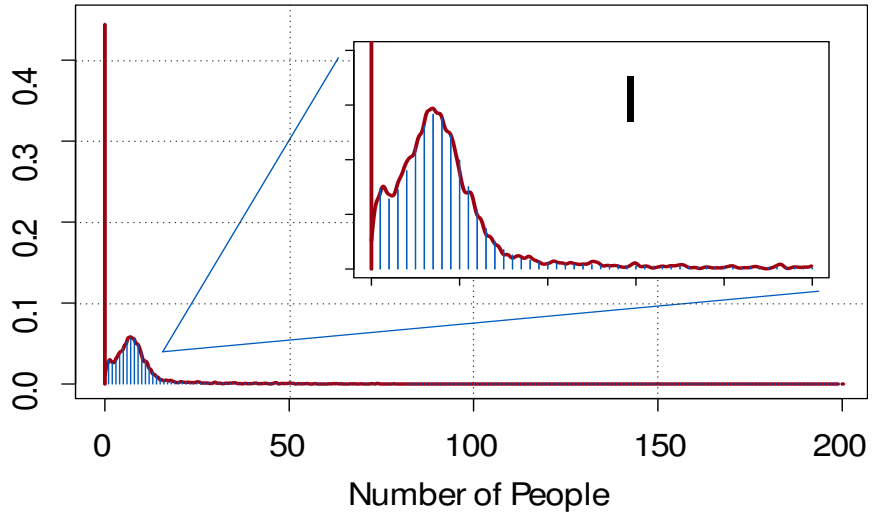
$$2A \xleftarrow{\mu} A + B \xrightarrow{\mu} 2B$$

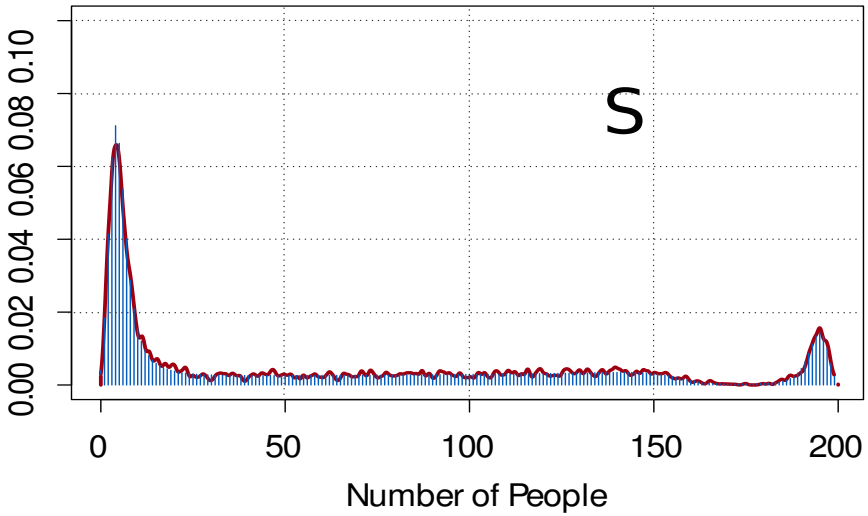
The deterministic approximation is very boring!
The diffusion approximation in such a case fails, only works up to the first hitting time of the boundary, and then it is singular.
At least three different approximations have been proposed to fix this problems (Schnoerr et al., Leite and Williams, Angius and B. and Horvath et al.), but still a mathematically sound approximation theory is to be found.

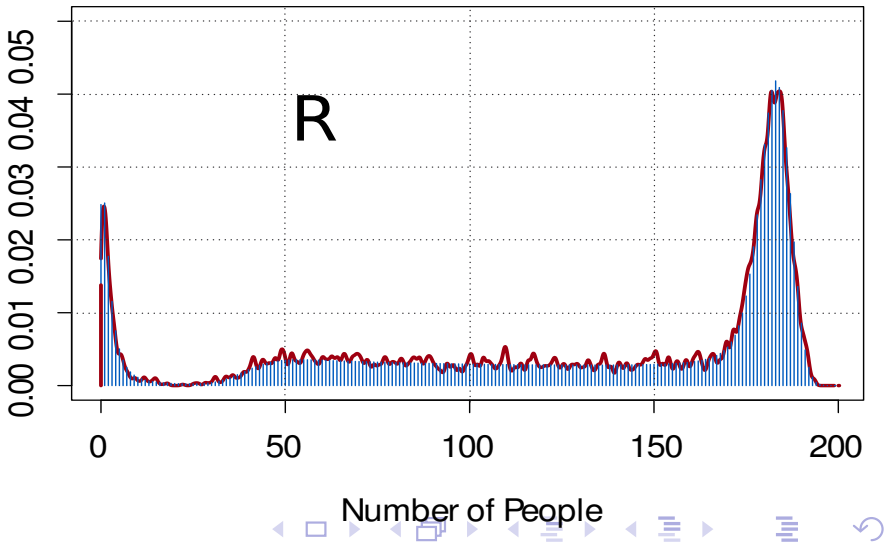
Another example in epidemiology (SIR)

The availbility of Infected makes a difference!









Relations with other topics of the project

- Fluid approximations are also used in the network models studied in topic T1
- Approximation theorems have been proved in some multi-scale setting, such as those of nested models in biomedicine, topic T2 of the project (next kick-off meeting next week)

Statistical Inference

The first problem to be addressed is parameter estimation. Let us keep in mind the following example

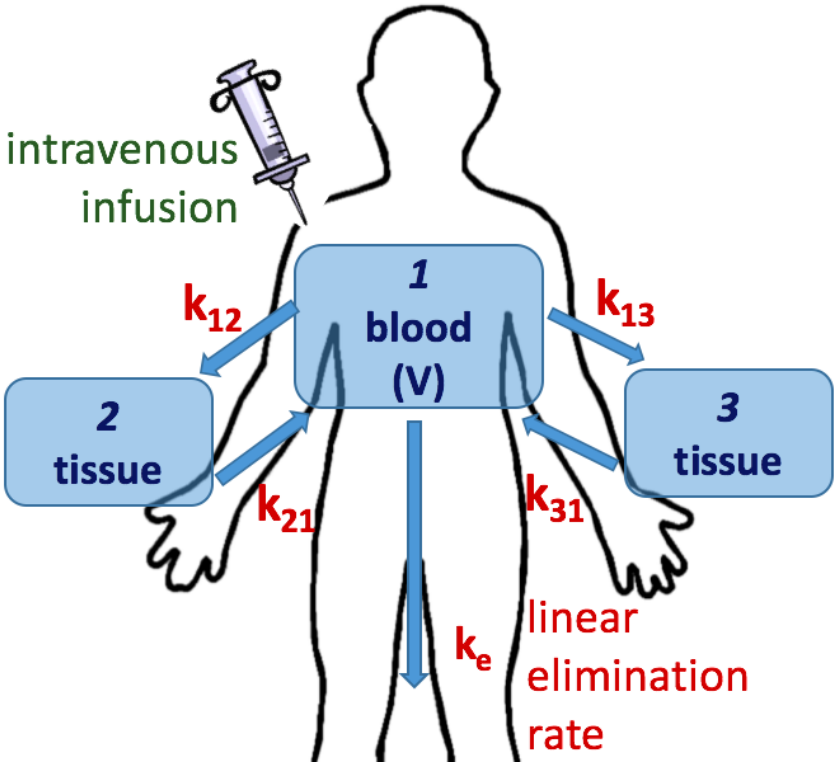
Compartmental models in parmacokinetics (PK)

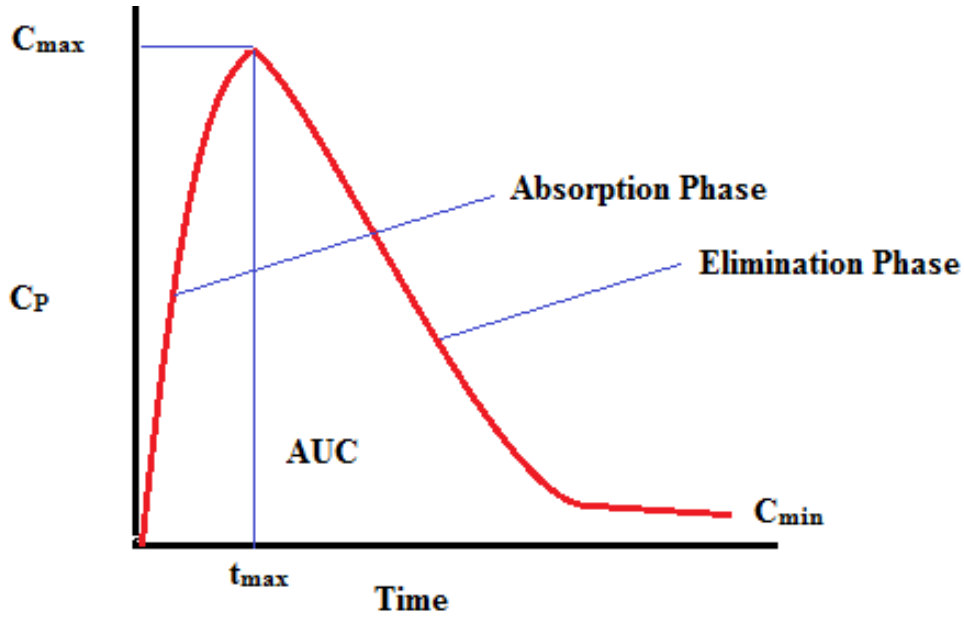
A drug is administered to a patient (e.g. by intravenous infusion). Molecules of such a drug are absorbed into the blood stream (B), possibly exchanged with some organ (O), and eliminated from the body. A reaction network for this model is

$$\emptyset \underset{\lambda_2}{\overset{\lambda_1}{\leftrightarrows}} B \underset{\lambda_4}{\overset{\lambda_3}{\leftrightarrows}} O$$

The system is most often modelled deterministically (cf. Gasparini, Erhardt, et al.), but diffusion models are considered as well.

Compartmental model





Typical obstacles to a straightforward inference

One or many of the following obstacles often appear in practice

- Data are observed discretely in time, maybe with a low frequency (e.g in PK blood is taken from patients only every few hours)
- Often the system is not fully observed: some components are latent. In PK, the concentration of the drug is usually only observed in the blood. The other compartments are not observed
- It is sometimes impossible to observe the system for large time windows (again in PK once the drug is eliminated there is no point in observing the same patient further). Non-trivial asymptotic schemes are needed.
- Different patients may carry different parameters $\rightarrow$ Random effects need to be included into the model
- Explicit expressions for the likelihood functions are often unknown or infeasible for numerical evaluation

Available technologies

Technologies for such problems are sometimes available but often need to be tailored to the specific problem, or extended to a different context, e.g.

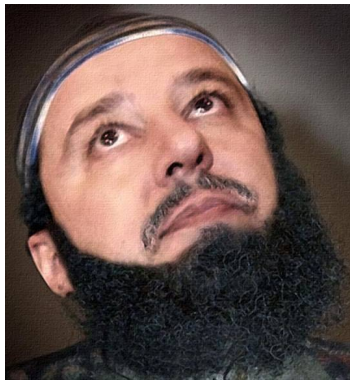
- EM algorithms have been invented to deal with unobserved components. Adaptations to stochastic process have been introduced (SAEM) and they need some quantities to be evaluated by Monte Carlo simulation. Fast algorithms are necessary.
- Random effects make inference numerically complicated even in the case of ODE modeling. Their inclusion into stochastic models have been attempted already, but further research is needed to make inference more practical.
- Bayesian strategies do exist, e.g. Approximate Bayesian Computations (ABC) but their applications in this context is non-trivial.
- Kalman and particle filters are also applied in this context

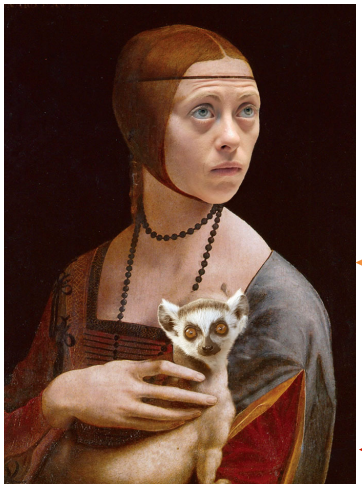
Approximated models

Since many of the methods for parameter estimations for stochastic processes requires simulations, the introduction of approximated models that are more easily simulated is very welcome.

E.g. diffusion approximations can be simulated much faster then Markov Chains.

This is also a reaction network ...





... but I do not want to comment it further