Simulation, Analysis, and Validation of Computational Models

— MC —

Lecturer: Michael Herrmann
School of Informatics, University of Edinburgh
michael.herrmann@ed.ac.uk, +44 131 6 517177

Overview

- Dynamical sytems
- Markow chains
- Monte Carlo

What is a simulation?

- Simulations are usually computer-based, using software models to provide support for decision makers and for training purposes. The purpose of the simulation is the imitation of the operation of processes in a real systems using a model of this system.
  - A model represents the key behaviours and characteristics of selected processes in a system
  - The simulation represents how the model evolves under different conditions over time.
- Simulation systems include
  - discrete event simulation,
  - process simulation, and
  - dynamic simulation,

  or a combination of all three across different subsystems.

adapted from https://www.twi-global.com/technical-knowledge/faqs/faq-what-is-simulation

States

- A state S_t describes a system at a particular time t.
- It contains all the available information needed to make solve a given optimisation or prediction task
- Example: a sequence 1, 2, 3, 4, 5, 6, ... that increases by +1 in every time step.
  - Given $S_1 = 1$, what is S_{11}?
  - Given $S_1 = 1$, what is S_{11}?
  - Given $S_1 = 1$, what is S_{11}?
- Example: S_{t+1} is usually the same as $S_t \in \{0, 1\}$, but at about every 10th step, $S_{t+1} = 1 - S_t$.
- Example: $S_{t+1} = \lambda S_t (1 - S_t)$ with $S_1 \in [0, 1]$ and $\lambda \in [0, 4]$.

Dynamical systems

Let $x \in \mathcal{S}$, a state space, and t_i in a set $\mathcal{I}$ of time points. For a given function

$$\Phi : \mathcal{I} \times S \to S,$$

a *dynamical system* is characterised by the two properties

- Identity
$$\Phi(0, x) = x \text{ and } 0 \in \mathcal{I}$$
- Semigroup property, i.e. if $t_2 + t_1 \in \mathcal{I}$
$$\Phi(t_2, \Phi(t_1, x)) = \Phi(t_2 + t_1, x)$$

$\Rightarrow$ The trajectory of the dynamical system:

$$\gamma_x \equiv \{\Phi(t, x) : t \in \mathcal{I}\}$$

Note: X and I can be discrete or continuous.

Example: Exponential growth

For $\mathcal{S} = (0, \infty)$, $\mathcal{I} = \{0, 1, 2, \dots\} = \mathbb{N}_0$, consider the dynamical system:

$$S_{t+1} = aS_t,$$

where $a > 0$ is a constant growth factor. Given S_0, we find

$$S_t = a^t S_0$$

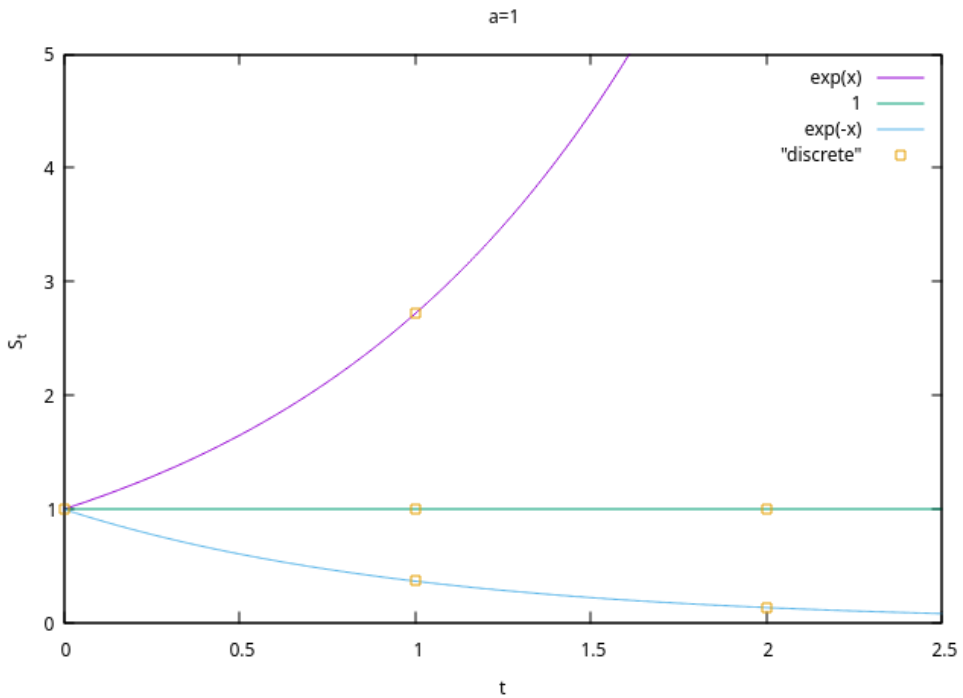
For a shifted versions of the sequence $\tilde{S}_0 = S_{t_1}$, we see

$$\tilde{S_{t_2}} = a^{t_2} \tilde{S}_0 = a^{t_2} a^{t_1} S_0 = a^{t_2+t_1} S_0 = S_{t_2+t_1}.$$

Asymptotic behaviour

$$\lim_{t\to\infty} S_t = \begin{cases} 0 & a < 1 \\ S_0 & a = 1 \\ \infty & a > 1 \end{cases}$$

Example: Exponential growth



Explicit system

$$S_t = a^t S_0$$

Asymptotic behaviour

$$\lim_{t\to\infty} S_t = \begin{cases} 0 & a < 1 \\ S_0 & a = 1 \\ \infty & a > 1 \end{cases}$$

Next time, we'll consider this example again, but in continuous time

Example: Exponential growth in 2D

$\mathcal{S} = (0, \infty)^2$, $\mathcal{I} = \{0, 1, 2, \dots\}$, consider the dynamical system:

$$\begin{pmatrix} S^1_{t+1} \\ S^2_{t+1} \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} S^1_t \\ S^2_t \end{pmatrix} = M \begin{pmatrix} S^1_t \\ S^2_t \end{pmatrix} = M\, S_t,$$

where $a > 0$ is a constant growth factor. Given S_0 we find

$$S_t = M^t S_0$$

If the sequence is shifted $\tilde{S}_0 = S_{t_1}$ then

$$\tilde{S_{t_2}} = M^{t_2} \tilde{S}_0 = M^{t_2} M^{t_1} S_0 = M^{t_2+t_1} S_0 = S_{t_2+t_1}.$$

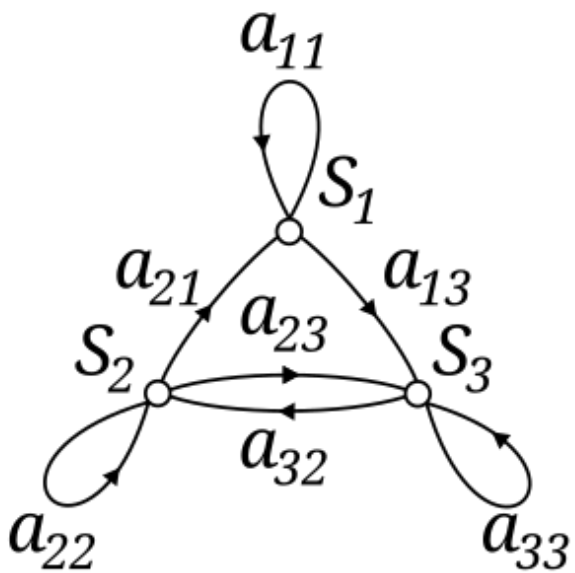
Asymptotic behaviour depends on the eigenvalues λ_i of M.

$$M = \sum_{i=1}^{2} \lambda_i \xi^i \xi^{i\top} = \sum_{i=1}^{2} \lambda_i \begin{pmatrix} \xi^i_1 \xi^i_1 & \xi^i_1 \xi^i_2 \\ \xi^i_2 \xi^i_1 & \xi^i_2 \xi^i_2 \end{pmatrix}, \text{ assuming } \lambda_i \in \mathbb{R}$$

if $\lambda_1 > \lambda_2$ and $\lambda_1 > 1$, divergence in the direction of eigenvector ξ_i (obviously, also a number of other cases is possible here).

Markov chains

Stochastic process described by
state probabilities and
state-transition probabilities



$$P(S_{t+1} = x_{j_{t+1}} \mid S_t = x_{j_t}, S_{t-1} = x_{j_{t-1}}, \dots) = P(S_{t+1} = x_{j_{t+1}} \mid S_t = x_{j_t})$$

Example: Weather Forecasting (see Li & Nakano, Ch. 2)

Weather forecasting over one hour, i.e. $\mathcal{I} = \{0, 1, 2, \dots\}$, given current state $S_t \in \{\text{sunny, cloudy, rainy}\}$:

	S_t =sunny	S_t =cloudy	S_t =rainy
S_{t+1} =sunny	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{2}{3}$
S_{t+1} =cloudy	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$
S_{t+1} =rainy	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$

Columns sum to 1.

Example: Weather Forecasting (see Li & Nakano, Ch. 2)

Markov property: Future depends on immediate past only.

If all we know about the state of the system is "$S_t = \text{sunny}$", then in $\frac{1}{3}$ of the cases, the weather will turn to "cloudy" in the next hour.

$$P(\text{sunny} \to \text{cloudy}) = P(S_{t+1} = \text{cloudy} | S_t = \text{sunny}) = \frac{1}{3}$$

$$\begin{pmatrix} P(S_{t+1} = \text{sunny}) \\ P(S_{t+1} = \text{cloudy}) \\ P(S_{t+1} = \text{rainy}) \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{6} \end{pmatrix} \begin{pmatrix} P(S_t = \text{sunny}) \\ P(S_t = \text{cloudy}) \\ P(S_t = \text{rainy}) \end{pmatrix}$$

$$\begin{pmatrix} \frac{1}{2} \\ \frac{1}{3} \\ \frac{1}{6} \end{pmatrix} = \begin{pmatrix} P(S_{t+1} = \text{sunny}) \\ P(S_{t+1} = \text{cloudy}) \\ P(S_{t+1} = \text{rainy}) \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{6} \end{pmatrix} \begin{pmatrix} P(S_t = \text{sunny}) = 1 \\ 0 \\ 0 \end{pmatrix}$$

Numerical Example: Weather Forecasting (Li & Nakano, Ch. 2)

See Notebook.

Perron Frobenius theorem

Theorem: The maximal eigenvalue of a matrix with positive entries is unique and has an eigenvector with positive entries.

If the matrix columns sum to 1, then this eigenvalue equals 1, and the respective eigenvector represents the stationary probabilities.

$$P_t^i = \sum_j M_{ij} P_t^j$$

Normalisation is conerved. If M is a stochastic matrix $\sum_i M_{ij} = 1$:

$$\sum_i P_{t+1}^i = \sum_i \sum_j M_{ij} P_t^j = \sum_j \underbrace{\left(\sum_i M_{ij} \right)}_{=1} P_t^j = \sum_j P_t^j$$

i.e. if $\sum_j P_t^j = 1$ then $\sum_j P_{t+1}^j = 1$.

Discussion

- For non-negative probability matrices, more assumptions are needed to show that M has a stationary distribution.
- Although more information should be available if the system runs over several time steps, Markov chains tend to reproduce only the average case for a system.
- Non-Markovianity is usually due to partial observability.
- If we are interested in the temporal variability, the stochastic matrix does not need to be the same in every step, but can be modulated by observations (see: Hidden Markov Models).

Monte Carlo method

- Buffon's needle problem (1733): Draw parallel lines with distance d, throw a needle of length ℓ, then the probability of an intersection is $p = \frac{2\ell}{\pi d}$, i.e. we can determine π by counting intersections.
- First discussed by Enrico Fermi (later analog "Electronic Numerical Integrator And Computer" was called FERMIAC,1947)
- The Monte Carlo method was developed to simulate nuclear reactions (Stanislaw Ulam, 1944)

Monte Carlo method

- Task is to calculate the expected value of a function f

$$E_X[f(X)] = \sum_{\text{all } x} P(x)\, f(x),$$

- If X is a high-dimensional random vector, this task may require exponentially many values.
- If eventually all the values "collapse" into an average, it may be sufficient to take samples of f based on some random x.
- Monte-Carlo methods become more reliable with methods to reduce variance and with suitable samplers.

Numerical Example: Monte Carlo (Li & Nakano, Ch. 1)

See Notebook.

Stratified sampling

- Stratified sampling (e.g. social strata in social science studies)
- Stratification should be *collectively exhaustive* and *mutually exclusive*
- A representative (or balanced) sample represents the strata proportionally
- For strata of same size N_h variance reduces by $(N_h/N)^2$.
- Potential problems: Sample should be large enough it represent all strata. Variance of properties can vary across strata. Trends may revert when combining strata (Simpson's paradox)

Importance sampling

- Variance can be reduced by sampling from the native distribution $p(X)$ but be a more convenient distribution $w(Y)$.
- Importance sampling re-weights the w-sample to provide information with respect to the original p distribution. Thus,

$$E[X] = \sum_{x \in \Omega} x p(x) = \sum_{x \in \Omega} \frac{x p(x)}{w(x)} w(x) = E\left[\frac{Y p(Y)}{w(Y)}\right]$$

- E.g. w can be a uniform distribution ("simple sampling")

$$\hat{E}_X[f(X)] = \frac{\sum_{y \in S} p(y) f(y)}{\sum_{y \in S} p(y)}$$

Metropolis algorithm

- New sample at $\boldsymbol{y} = \boldsymbol{x} + \rho \boldsymbol{q}$, where ρ is the search radius and $\boldsymbol{q}$ is a random vector of ± 1.
- The new point $\boldsymbol{y}$ is accepted if it is better than the previous point, or, if not, then only with a probability $p_A = \exp\left(-\frac{\Delta E}{kT}\right)$, where k is a constant, T is an adaptable temperature parameter, and ΔE is the loss in quality.
- The quality E is e.g. the relative number of errors or is implied by the physics of the problem.
- Metropolis-Hastings: $p_{\mathrm{A}} = \min\left(1, \frac{p(\boldsymbol{x} \mid \boldsymbol{y}) w(\boldsymbol{y})}{p(\boldsymbol{y} \mid \boldsymbol{x}) w(\boldsymbol{x})}\right)$ works for an arbitrary sampling distribution w (see importance sampling, slide 19).

Markov-Chain-Monte-Carlo (MCMC)

- Construct a Markov chain that with a given stationary distribution
- It is still a problem to find out how long it takes to reach this stationary distribution.
- Samples have usually a high auto-correlation (i.e. are similar), which can lead to a reduction of the variance, but also to a bias (underestimation) of the error of the target quantity (see slide 16)

Conclusion

- Systems can be understood by making use of known facts or by exemplary simulation.
- The Markov property, independence assumptions, reachability, sufficiently large sample size and others need to be critically checked each in each case.
- None of the points made here is as such important for this course, but we may later need some of them again. Then please check back here.

Next topics

- Quantitative description of continuous systems
- Linear systems
- Stochastic systems
- Iterated functions