Optimization and Simulation

Introduction to simulation

Michel Bierlaire

Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne

Outline

Modeling

System

- A system can be seen as a black box, modeled by

$$z = h(x, y, u)$$

- Example: a car
- x captures the state of the system (e.g. speed, position of other vehicles)
- y captures external influences (e.g. wind)
- u captures possible human controls on the system (e.g. acceleration/deceleration)
- z represents indicators of performance (e.g. oil consumption).

Modeling

Decompose the complexity

- The model h is usually decomposed to reflect the interactions of the subsystems
- For example,
  - a car-following model captures the target speed of the driver,
  - an engine model derives the actual consumption as a function of the acceleration.

Simulation

- Captures the causal effects.
- Captures the uncertainty.

Simulation

Definition

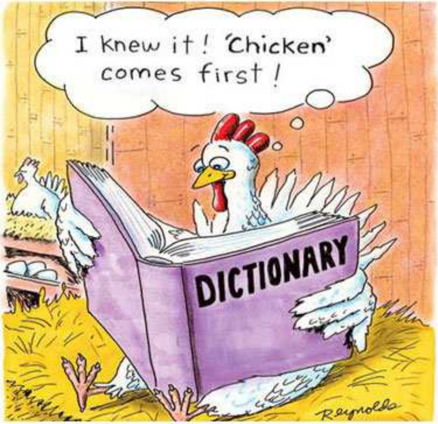
the act of imitating the behavior of some situation or some process by means of something suitably analogous

Outline

Modeling

Causal effects

- Very important to identify the causal effects
- Failure to do so may generate wrong conclusions



Example: improving safety

Accidents in Kid City

- The mayor of Kid City has commissioned a consulting company
- Objective: assess the effectiveness of safety campaigns
- They propose to use simulation

Example: improving safety

Accidents in Kid City

Example: improving safety

Accidents in Kid City:

Example: improving safety

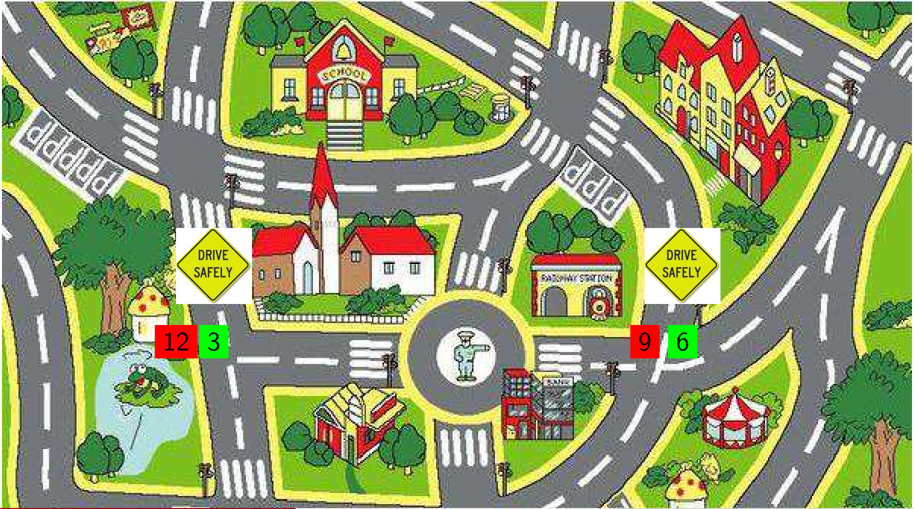
Accidents in Kid City

Example: improving safety

Accidents in Kid City

Example: improving safety

Accidents in Kid City:



Example: improving safety

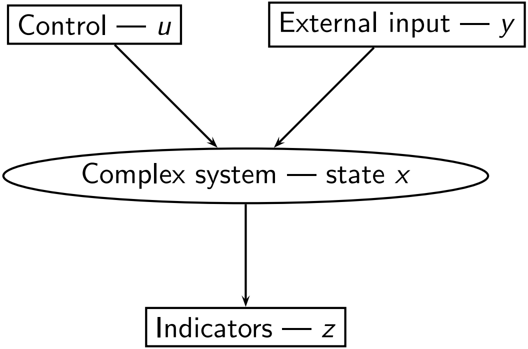
Two major flaws

- Causal effects are not modeled
- Simulation performed with only one draw

Simulation: what it is not

Simulation: what it is not

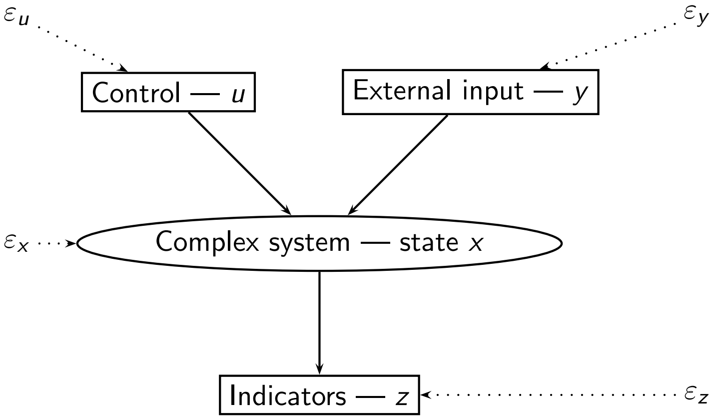
$$z = h(x, y, u)$$



Outline

Simulation

$$Z = h(X, Y, U) + \varepsilon_z$$



Simulation

Propagation of uncertainty

$$Z = h(X, Y, U) + \varepsilon_z$$

- Given the distribution of X, Y, U and ε_z
- what is the distribution of Z?

Derivation of indicators

- Mean
- Variance
- Modes
- Quantiles

Simulation

Sampling

- Draw realizations of X, Y, U, ε_z
- Call them $x^r, y^r, u^r, \varepsilon_z^r$
- For each r, compute

$$z^r = h(x^r, y^r, u^r) + \varepsilon_z^r$$

- z^r are draws from the random variable Z

Statistics

"Numbers don't lie. That's where we come in."

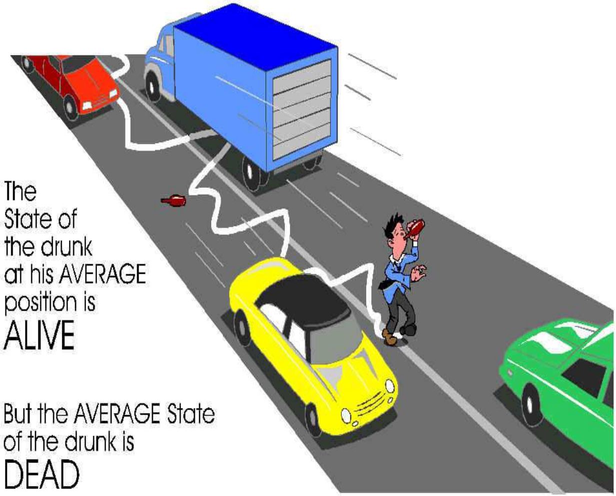
Indicators

- Mean: $\mathrm{E}[Z] \approx \bar{Z}_R = \frac{1}{R}\sum_{r=1}^{R} z^r$
- Variance: $\mathrm{Var}(Z) \approx \frac{1}{R}\sum_{r=1}^{R}(z^r - \bar{Z}_R)^2$.
- Modes: based on the histogram
- Quantiles: sort and select

Important: there is more than the mean

Outline

The mean



[Savage et al., 2012]

The mean

The flaw of averages

[Savage et al., 2012]

$$\mathrm{E}[Z] = \mathrm{E}[h(X, Y, U) + \varepsilon_z] \neq h(\mathrm{E}[X], \mathrm{E}[Y], \mathrm{E}[U]) + \mathrm{E}[\varepsilon_z]$$

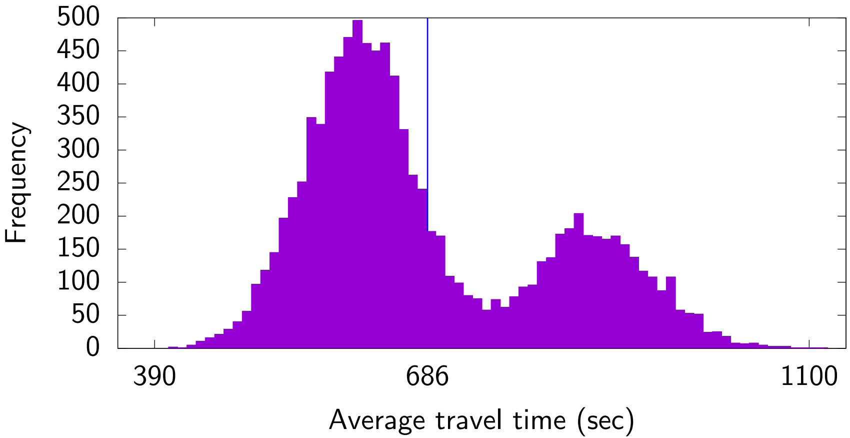
... except if h is linear.

There is more than the mean

Example

- Intersection with capacity 2000 veh/hour
- Traffic light: 30 sec green / 30 sec red
- Constant arrival rate: 2000 veh/hour during 30 minutes
- With 30% probability, capacity at 80%.
- Indicator: Average time spent by travelers

There is more than the mean



Outline

Pitfalls of simulation

Few number of runs

- Run time is prohibitive
- Tempting to generate partial results rather than no result

Focus on the mean

- The mean is useful, but not sufficient.
- For complex distributions, it may be misleading.
- Intuition from normal distribution (mode = mean, symmetry) do not hold in general.
- Important to investigate the whole distribution.
- Simulation allows to do it easily.

Challenges

- How to generate draws from Z?
- How to represent complex systems? (specification of h)
- How large R should be?
- How good is the approximation?

Pseudo-random numbers

Definition

- Deterministic sequence of numbers
- which have the appearance of draws from a $U(0,1)$ distribution

Typical sequence

$$x_n = ax_{n-1} \text{ modulo } m$$

- This has a period of the order of m
- So, m should be a large prime number
- For instance: $m = 2^{31} - 1$ and $a = 7^5$
- x_n/m lies in the $[0,1[$ interval

Outline of the lectures

- Drawing from distributions
- Discrete event simulation
- Data analysis
- Variance reduction
- Markov Chain Monte Carlo

Reference

[Ross, 2006]

Bibliography

Ross, S. M. (2006).
Simulation.
Elsevier, fourth edition.

Savage, S., Danziger, J., and Markowitz, H. (2012).
The Flaw of Averages: Why We Underestimate Risk in the Face of Uncertainty.
Wiley.