ISE/OR 762
Stochastic Simulation Techniques

Topic 0: Introduction to Discrete Event Simulation

Yunan Liu

Department of Industrial and Systems Engineering
NC State University

January 9, 2018

Why Simulation?

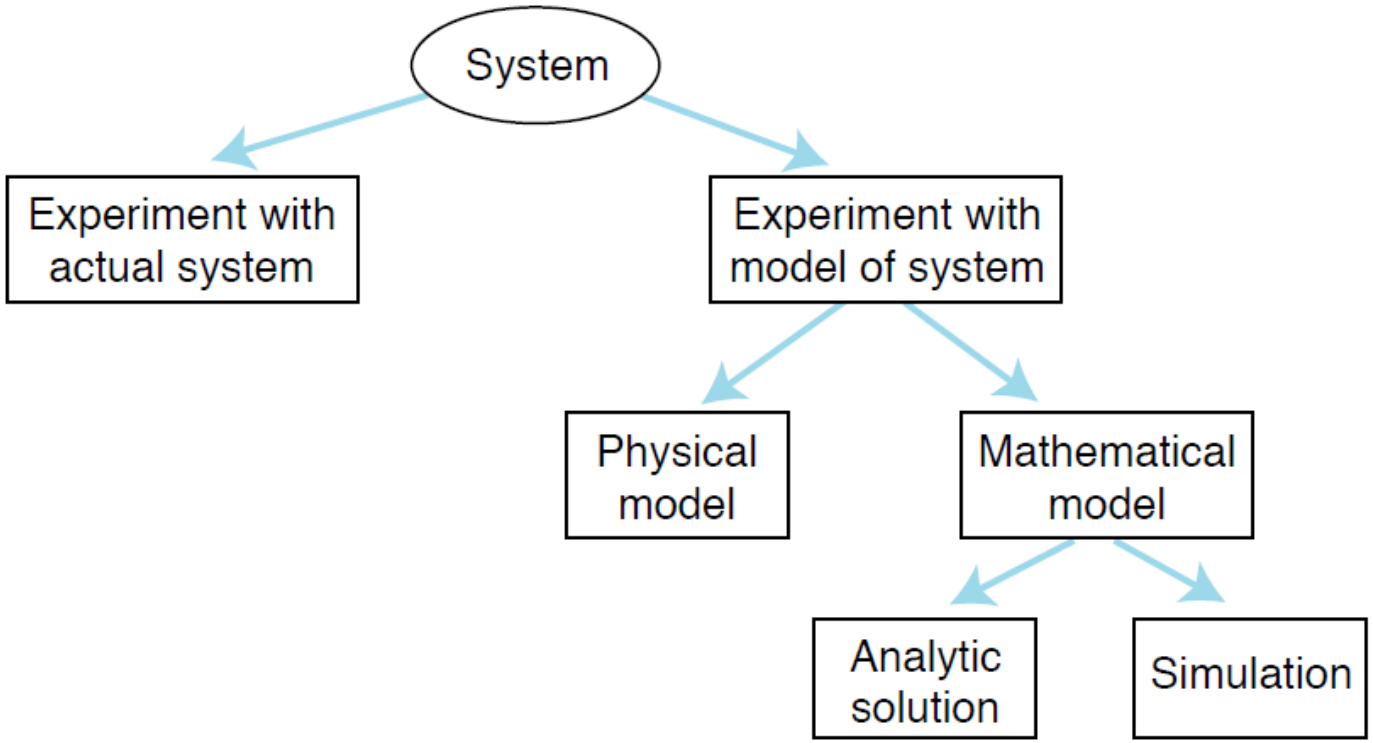
- Simulation is used to "simulate" the operations of various kinds of real-world systems or processes with the aid of a computer.
- We make assumptions about how a system works; these assumptions constitute a model.
- Models are used to gain some understanding of how a system behaves.
- Analysis approach based on system complexity:
  - Simple models: use analytical (mathematical) methods to obtain exact information on questions of interest.
  - Complex models: use simulation to evaluate the model numerically.

System Analysis

System: a collection of entities, e.g. people or machines, that act and interact together toward the accomplishment of some logical end.

State: collection of variables necessary to describe a system at a particular time, relative to the objectives of a study.

A system can be *discrete* or *continuous* depending on how the state variables change with respect to time.



Application of Simulation

- Designing and analyzing manufacturing and inventory systems
- Determining hardware requirements or protocols for communication networks
- Determining hardware and software requirements for computer systems
- Designing and operating transportation systems such as airports, freeways, sports and subways
- Evaluating designs for service organizations such as call centers, fast-food restaurants, hospitals and post offices
- Analyzing financial or economic systems; pricing complex financial instruments

Different Kinds of Simulation

- Static v.s. dynamic
  - Does time have a role in the model?
- Continuous-change v.s. discrete-change
  - Can "state" change continuously, or only a discrete points in time?
- Deterministic v.s. stochastic
  - Is everything for sure or is there uncertainty?
- By hand v.s. by computer
  - Is the model complicated so we need a computer?
- Most operational models:
  - Dynamic, discrete-change, stochastic, computer-based

Discrete Event Simulation

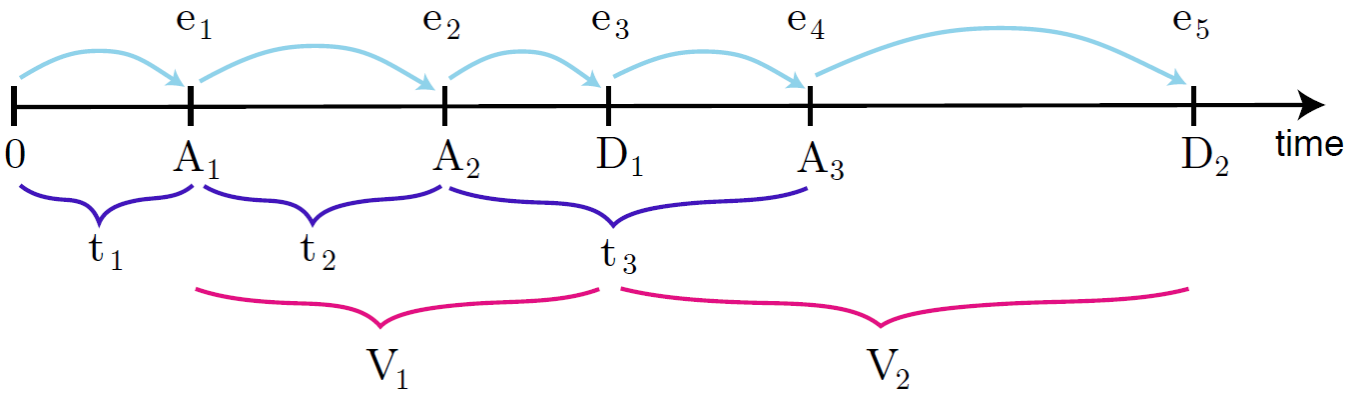
Objective: Model a system as it evolves over time by a representation in which the state variables change instantaneously at separate points in time, e.g. a queue.

Characteristics:

- Keep track of the simulated (random) time as simulation proceeds.
- Advance simulated time via a "simulation clock".
- Simulation clock can be advanced either when an event occurs or at a fixed increment of time.
- Sample system states as time evolves to prepare for output estimators.

Example: A Single-Server Queue

- Random elements:
  - A_i = time of arrival of the ith customer;
  - $t_i = A_i - A_{i-1}$ = interarrival time ($t_0 = 0$);
  - V_i = processing (service) time of ith customer;
  - W_i = waiting time of ith customer (excluding service);
  - $D_i = A_i + W_i + V_i$ = departure time of ith customer.
- Possible events: arrivals and departures.
- Advance simulation clock when an event occurs
  - e_i = time of the ith event.
- V_i's and t_i's distributed according to a known distribution.



Estimating Useful Performance Metrics

- Estimate the average waiting time of customers

$$N(t) = \text{ number of customers that have arrived in } [0, t]$$

$$\widehat{W}(T) = \frac{1}{N(T)} \sum_{i=1}^{N(T)} W_i$$

- Estimate the average number of customers in the system

$$Q(t) = \text{ total number of customers in the system at time t}$$

$$\widehat{Q}(T) = \frac{1}{T} \int_0^T Q(t)dt.$$

- Estimate the average utilization

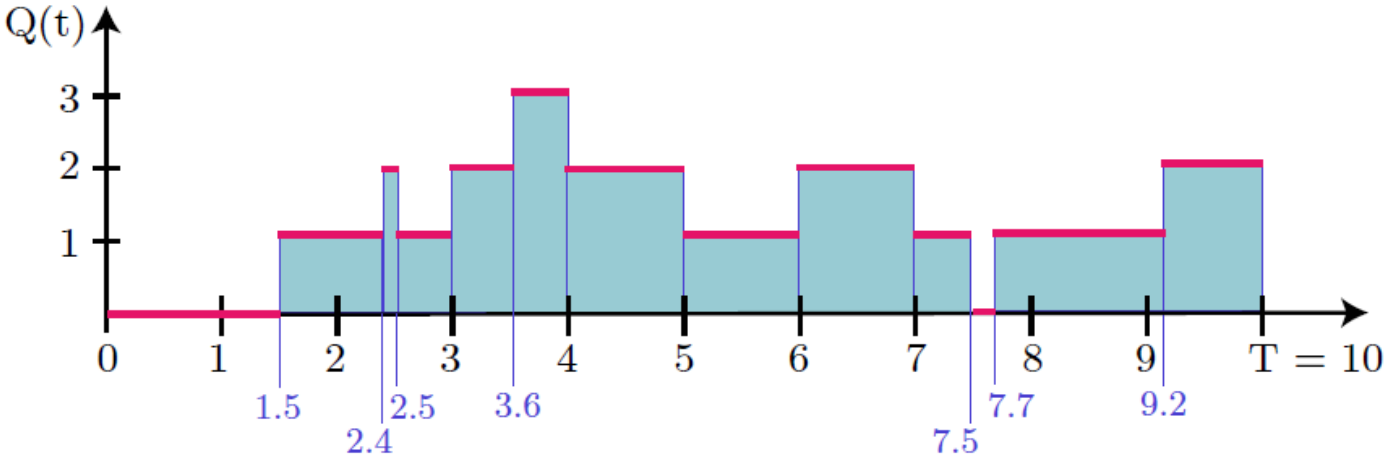
$$\widehat{\rho}(T) = \frac{1}{T} \int_0^T B(t)dt, \qquad B(t) = \mathbf{1}_{\{Q(t)>0\}}.$$

An Example with Numbers

- Set $T = 10$
- Initialize system so that there are no customers at time 0
- Generate arrivals and service times according to their distributions
  Arrival times: 1.5, 2.4, 3, 3.6, 6, 7.7, 9.2, 11
  Processing times: 1, 1.5, 1, 2, 0.5, 3.5, 1.5, 1

Time	Event	Next event (type & time)
1.5	A_1	$A_2 = 2.4$
2.4	A_2	$D_1 = 2.5$
2.5	D_1	$A_3 = 3$
3	A_3	$A_4 = 3.6$
3.6	A_4	$D_2 = 4$
4	D_2	$D_3 = 5$
5	D_3	$A_5 = 6$
6	A_5	$D_4 = 7$
7	D_4	$D_5 = 7.5$
7.5	D_5	$A_6 = 7.7$
7.7	A_6	$A_7 = 9.2$
9.2	A_7	

Simulation Outputs



Compute desired performance functions:

- $N(T) = 7$
- $\widehat{W}(T) = \frac{0+0.1+1+1.4+1+2}{7} = 0.7857$
- $\widehat{Q}(T) = \frac{1\times(.9+.5+1+.5+1.5)+2\times(.1+.6+1+1+.8)+3\times.4}{10} = 1.26$
- $\widehat{\rho}(T) = \frac{1}{10}\left[10 - 1.5 - .2\right] = 0.83$

Monte Carlo Simulation

- Definition: Computational algorithms that rely on repeated random sampling to obtain numerical results. (A simulation estimator based on strong law of large numbers.)
- Named after the gambling hot spot in Monaco: chance and random outcomes are central to the modeling technique, much as they are to games like roulette, dice and slot machines.
- First developed by Stanislaw Ulam and John Von Neumann.
- Applications: Evaluate complicated integrals (multiple dimensions), pricing complex financial instruments (options)
- Example: Estimate the value of

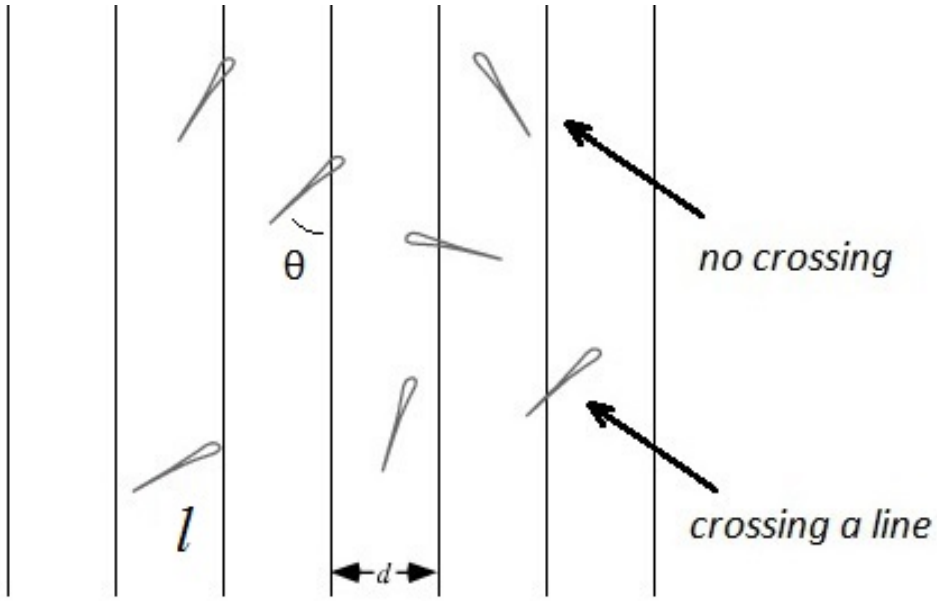
$$I \equiv \int_a^b g(x)dx.$$

Generate $X \sim \text{Unif}(a, b)$, set $Y = (b-a)g(X)$. Then

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \approx \mathbb{E}[Y] = (b-a)\mathbb{E}[g(X)] = (b-a)\int_a^b g(x)\,\frac{1}{b-a}dx = \int_a^b g(x)dx.$$

Buffon's Needle Problem: Precomputer Simulation in 1777

Buffon is remembered in the history of probability theory for his famous needle problem - the first example of a simulation experiment



Buffon's Needle Problem:

If a floor has equally spaced parallel lines a distance d apart and if a needle of length l is tossed at random on the floor where $l \leq d$, then what is the probability that the needle will intersect a line?

Buffon's Needle Problem: Precomputer Simulation in 1777

The distance from the head of a needle to a line is $X \sim \text{Unif}(0, d)$.

Conditioning on Θ:

$$\mathbb{P}\left(\text{Needle intersects a line}|\Theta = \theta\right) = \mathbb{P}\left(X \in [0, l\sin\theta]\right) = \int_0^{l\sin\theta} \frac{1}{d}\mathrm{d}x = \frac{\sin\theta \cdot l}{d}.$$

Unconditioning on Θ yields:

$$\begin{aligned}\mathbb{P}\left(\text{Needle intersects a line}\right) &= \int_0^{\frac{\pi}{2}} \mathbb{P}\left(\text{Needle intersects a line}|\Theta = \theta\right) \cdot \frac{1}{\pi/2}\mathrm{d}\theta \\ &= \int_0^{\frac{\pi}{2}} \frac{\sin\theta \cdot l}{d} \cdot \frac{2}{\pi}\mathrm{d}\theta = \frac{2l}{\pi d} = 0.637.\end{aligned}$$

When $l = d$, $\mathbb{P}\left(\text{Needle intersects a line}\right) = \frac{2}{\pi} = 0.637$.

Buffon's Needle Problem: Precomputer Simulation in 1777

Toss a needle for a large number N times. Define independent r.v.'s:

$$X_i = \mathbf{1}_{\{i^{\text{th}} \text{ needle crosses a line}\}} = \begin{cases} 1, & \text{if } i^{\text{th}} \text{ needle crosses a line;} \\ 0, & \text{otherwise.} \end{cases}$$

According to the strong law of large number (SLLN):

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} X_i = \frac{2}{\pi} \qquad \text{so that} \qquad \pi \approx \frac{2}{\frac{1}{N}\sum_{i=1}^{N} X_i} \quad \text{for } N \text{ large.}$$

Remarks:

- Buffon's needle-tossing experiment can be conducted by hand.
- It is the earliest example of using independent replications of a simulation to approximate an important physical constant.

Simulations in TV Shows

British Sci-Fi TV drama on Netflix: **Black Mirror**
Season 4, Episode 4: "Hang the DJ"

- An algorithm simulating the dating outcomes of 1000 independent digital copies of the consciousness of Frank and Amy.
- The app promises a 99.8% percent "perfect" match.

Issues Arising in Monte Carlo Simulation

- How to generate needed input random objects?
- How many computer experiments do we need?
- How to compute expectations associated with limit stationary distributions?
- How to exploit problem structure to speed up the computation?
- How to efficiently compute probabilities of rare events?
- How to estimate sensitivity of a stochastic model to changes in a parameter?
- How to use simulation to optimize our choice of decision parameters?

Advantages and Disadvantages of Simulation

- Advantages:
  - Simulation allows to estimate the performance of complex, real-world, systems that cannot be evaluated analytically.
  - Allows us to study long time frames in compressed time.
  - Provides better control of experimental conditions.
- Disadvantages:
  - Lacking of insights.
  - Expensive and time-consuming to develop.
  - Produces only estimates (not exact) of a model's true characteristics.
    - Implementation errors;
    - Sampling errors;
    - Machine errors;
    - Statistical errors.

Pitfalls of Simulation

- Lack of well defined objectives.
- Failure to collect good system data.
- Failure to account correctly for sources of randomness.
- Use of arbitrary distributions.
- Assume independence when it is not present.
- Not running enough simulations.
- Treating simulation results as "true answers".
- Replacing random variables by their means.

Danger of Replacing Random Variables by Their Means

Example: Consider a manufacturing system consisting of a single machine.

- "Raw" parts arrive according to a Poisson process with rate $\lambda = 1$ per min.
- Processing times are exponentially distributed with mean $1/\mu = 0.99$ min.

This is an $M/M/1$ queue with traffic intensity $\rho = 0.99$ (ISE760)

- The long-run average number of customers in the system

$$\mathbb{E}[Q(\infty)] = \lim_{t\to\infty} \frac{1}{t}\int_0^t Q(u)du = \frac{\rho}{1-\rho} = 99.$$

- The long-run average waiting time

$$\mathbb{E}[W_\infty] = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} W_i = \frac{1}{\mu}\frac{\rho}{1-\rho} = 98.01 \text{ minutes.}$$

However, if we replace both distributions by their corresponding means, then

- The average number of customers in system ≈ 1;
- No part is ever delayed in the queue!

Conclusion: Ignoring randomness may result in extremely inaccurate estimates.

Rest of the Class . . .

- Review of probability theory (Topic 1).
- How to generate the random elements (variable, process, permutation, etc.) following various distributions (Topics 2–3).
- How to use Monte Carlo simulation to price path-dependent financial instruments (Topic 4).
- How to use the generated random elements to characterize the dynamics of a realistic system (Topic 5).
- Given a realization of a system, how to design effective procedures to estimate desired performance measures (Topic 6).
- When many sample paths (realizations) are needed, how to design efficient method to reduce number of samples (shortening the total computation time, Topic 7).
- How to select appropriate probability distributions (and their parameters) for the simulation algorithms (Topic 8).