Simulation

CS 700

1

Acknowledgement

These slides are based on presentations created and copyrighted by Prof. Daniel Menasce (GMU)

2

Types of simulation

- Emulation
  - Simulation using hardware or firmware, e.g. execute IA32 (Intel) programs on Power-PC platforms by emulating instructions
- Monte-Carlo simulation
  - No notion of time
  - Used to model probabilistic phenomenon that do not change characteristics with time
- Trace-driven simulations
  - Inputs are event traces collected from a real system
- Discrete-event simulation
  - Uses a discrete-model of the system being simulated

3

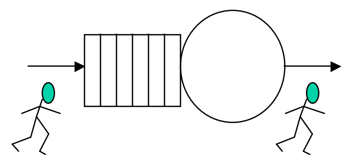
Components of a Simulation Model

- Event Generation:
  - Trace-driven
  - Distribution-driven
  - Hybrid
- Event Processing
  - Calendar of Events
  - Event-handling procedures
- Transaction List (with parameters)
- Queues
- Simulation Clock
- Computation of Statistics

4

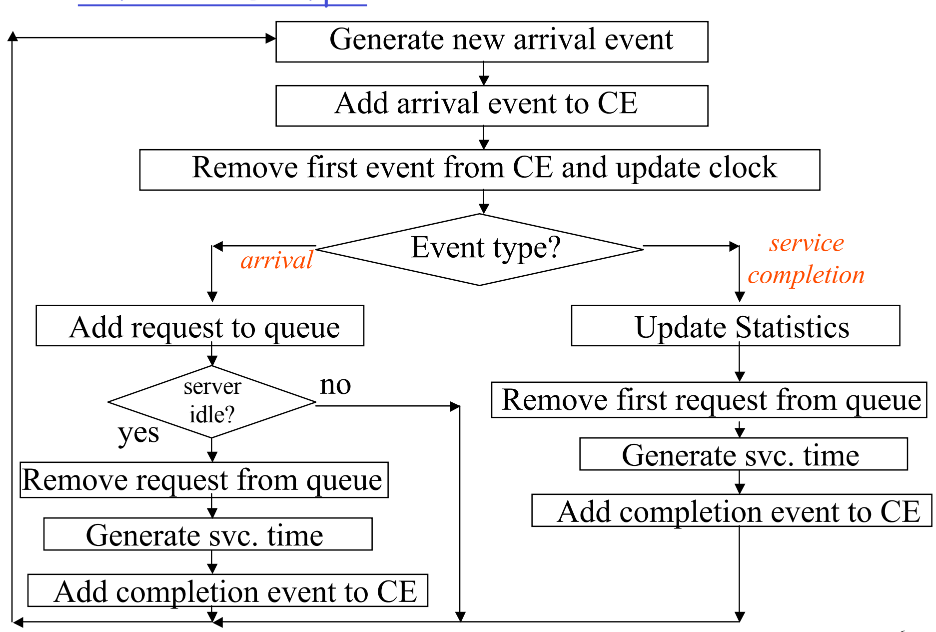
Discrete-event Simulation Example: Single Queue

- Events:
  - Arrival of a customer
  - Service completion
- Statistics:
  - Total number of arrivals
  - Total departures
  - Total server busy time
  - Total waiting time
  - Total departures from queue
  - Total squares of waiting time



5

Simulation Example



6

Calendar of Events

Event Type	Event Time	Event Parameters
arrival	10.5	…..
arrival	12.8	…..
completion	13.1	…..
…	…	…

- The calendar of events is ordered in increasing chronological order.
- Parameters may include the transaction Id associated with the event.

7

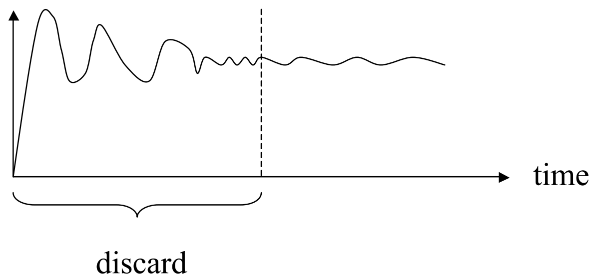
Common Mistakes in Simulation

- Inappropriate level of detail:
  - Too detailed: more development time and higher likelihood of bugs
  - Should start with a less detailed model first and increase complexity as needed.
- Unverified Models:
  - Simulation programs are usually large and complex programs and may have bugs that invalidate the results.
- Invalid Models:
  - Incorrect assumptions may be used. Need to validate through analytic models, measurements, and or intuition.

8

Common Mistakes in Simulation

- Improperly Handled Initial Conditions:
  - Should discard first part of run: transient behavior.



9

Common Mistakes in Simulation

- Improper simulation length.
- Poor Random Number Generator.
- Improper Selection of Seeds.

10

Verifying Simulation Models

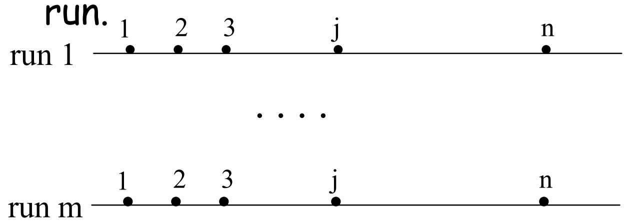
- Trace Analysis: examine traces of a few transactions as they go through the system.
- Continuity Test: small variations in the input should show small variations in the output.
- Check Extreme Values: extreme values (e.g., low loads or very high loads) should be easy to verify by crude analytic models.

Verifying Simulation Models

- Check for Basic Relationships: verify if results satisfy basic laws (e.g., Little's Law).
- Bound validation: use, if possible, existing analytic models for situations that are known to be upper or lower bounds
- Trend verification: check if the trends shown by the model match your intuition.
- Numeric range validation: check if the numerical results are within expected numerical ranges.

Transient Elimination with Independent Runs

- Run *m* runs of the simulation with a different seed for each run.
- Each run has *n* observations.
- Let $x_{i,j}$ be the j-th observation in the i-th run.



Transient Elimination with Independent Runs

Step 1: compute average of j-th observation over all runs.

$$\bar{x}_j = \frac{1}{m}\sum_{i=1}^{m} x_{i,j}$$

Step 2: compute the overall average.

$$\bar{\bar{x}} = \frac{1}{n}\sum_{j=1}^{n} \bar{x}_j$$

Step 3: Set the number of deleted observation, k, equal to 1.

Step 4: Compute the overall mean without the first k observations.

$$\bar{\bar{x}}_k = \frac{1}{n-k}\sum_{j=k+1}^{n} \bar{x}_j$$

Transient Elimination with Independent Runs

Step 5: compute the relative change Δ

$$\Delta = \frac{\bar{\bar{x}}_k - \bar{\bar{x}}}{\bar{\bar{x}}}$$

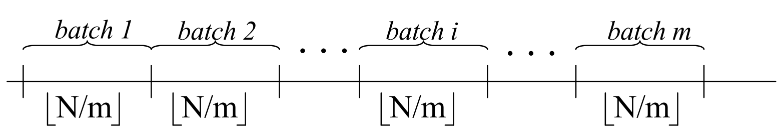
Step 6: If $|\Delta_\kappa - \Delta_{\kappa-1}| >$ threshold then do $k \leftarrow k + 1$ and go to step 4.

Step 7: Remove the first k observations and use $\bar{\bar{x}}_k$ as the average.

15

Transient Elimination with Batch Means

- Single run with N observations.
- Divide the run into *m* sub-samples called batches of size $n = \lfloor N/m \rfloor$.
- Let $x_{i,j}$ be the j-th observation in the i-th batch.



16

Transient Elimination with Batch Means

Step 1: Set n = 2

Step 2: compute the average of the i-th batch.

$$\bar{x}_i = \frac{1}{n}\sum_{j=1}^{n} x_{i,j}$$

Step 3: compute the overall average.

$$\bar{\bar{x}} = \frac{1}{m}\sum_{i=1}^{m} \bar{x}_i$$

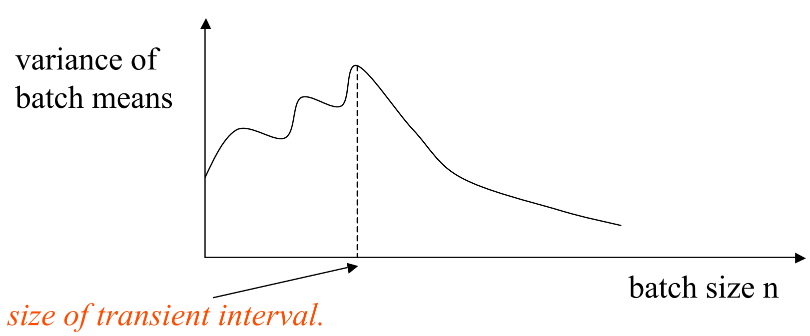
Step 4: Compute the variance of the batch means:

$$Var(\bar{x}) = \frac{1}{m-1}\sum_{i=1}^{m} (\bar{x}_i - \bar{\bar{x}})^2$$

17

Transient Elimination with Batch Means

Step 5: Increase n by 1 and repeat steps 2-4 and plot the variance as a function of n. The point at which the variance starts to decreases is the length of the transient interval.



18

Stopping Criteria: Independent Runs

- Run *m* runs of the simulation with a different seed for each run.
- Each run has $n + n_o$ observations where n_o is the size of the transient phase.
- The number *n* is increased until the precision in the confidence interval reaches a desired value.

19

Stopping Criteria
Independent Runs

Step 0: Initialization: n = 100.

Step 1: compute the mean for each replication.

$$\bar{x}_i = \frac{1}{n} \sum_{j=n_0+1}^{n} x_{i,j}$$

20

Stopping Criteria
Independent Runs

Step 2: compute the overall mean for all replications.

$$\bar{\bar{x}} = \frac{1}{m}\sum_{i=1}^{m} \bar{x}_i$$

Step 3: compute the variance of the replicate means.

$$\mathrm{Var}(\bar{x}) = \frac{1}{m-1}\sum_{i=1}^{m} (\bar{x}_i - \bar{\bar{x}})^2$$

Step 4: compute the confidence interval for the mean as:

$$\bar{\bar{x}} \pm t_{[1-\alpha/2,m-1]} \frac{\sqrt{\mathrm{Var}(\bar{x})}}{\sqrt{m}}$$

21

Stopping Criteria - Independent Runs

Step 5: compute the accuracy r as.

$$r = \frac{\left(t_{[1-\alpha/2,m-1]} \frac{\sqrt{\mathrm{Var}(\bar{x})}}{\sqrt{m}} \right)}{\bar{\bar{x}}} \times 100$$

Step 6: If r > desired value (e.g., 5) then n = n + 100 and go to Step 1, else STOP.

22

Stopping Criteria: Independent Runs

- Number of discarded observations: $m \times n_o$
- To reduce the number of wasted observations use a small value of *m*.

23

Stopping Criteria: Batch Means

- Single run with $N + n_o$ observations where n_o is the size of the transient phase.

Step 0: Start with a small value of n (e.g., 1).
Step 1: compute the mean for each batch.

$$\bar{x}_i = \frac{1}{n}\sum_{j=1}^{n} x_{i,j}$$

24

Stopping Criteria
Batch Means

Step 2: compute the overall mean for all batches.

$$\overline{\overline{x}} = \frac{1}{m}\sum_{i=1}^{m}\overline{x}_i$$

Step 3: compute the variance of the batch means.

$$\mathrm{Var}(\overline{x}) = \frac{1}{m-1}\sum_{i=1}^{m}(\overline{x}_i - \overline{\overline{x}})^2$$

Step 4: compute the confidence interval for the mean as:

$$\overline{\overline{x}} \pm t_{[1-\alpha/2;m]}\sqrt{\frac{\mathrm{Var}(\overline{x})}{m}}$$

25

Stopping Criteria
Batch Means

Step 5: compute the auto-covariance

$$\mathrm{Cov}(\overline{x}_i, \overline{x}_{i+1}) = \frac{1}{m-2}\sum_{i=1}^{m-1}(\overline{x}_i - \overline{\overline{x}})(\overline{x}_{i+1} - \overline{\overline{x}})$$

Step 6: Check for proper batch size: If $\mathrm{Cov}(\overline{x}_i, \overline{x}_{i+1}) << \mathrm{Var}(\overline{x})$ then stop. Otherwise, double n and go to step 1.

26

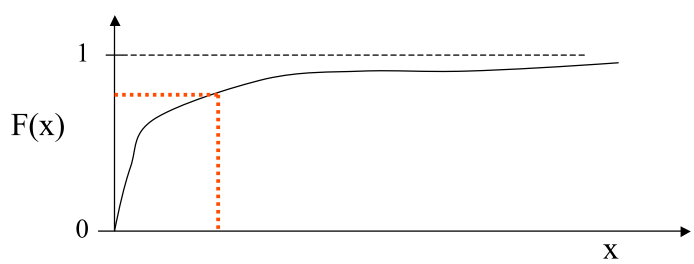
Seed Selection

- Never use zero as a seed.
- Avoid even values.
- Reuse seed for repeatability of experiments.
- Do not use random seeds (e.g., system time) if the simulation is to be repeated.

27

Generation of Random Variables

- Assume that *u* is a value uniformly distributed between 0 and 1.
- Method of the inverse of the CDF:



28

Generation of Random Variables

- Assume that *u* is a value uniformly distributed between 0 and 1.
- CDF for the exponential: $1 - e^{-x/a}$
  - Inverse of the CDF: -a ln(u)
- CDF for the Pareto distribution: $1 - x^{-a}$
  - Inverse of the CDF: $1/u^{1/a}$

29