COMP 190-088: Systems Performance Analysis

Analysis of Simulation Results

Jasleen Kaur

Department of Computer Science
The University of North Carolina at Chapel Hill

Spring 2005

Simulations: Outline

- Simulation Basics
  - Terminology
  - Types of simulations
  - Implementing a simulator
- **Analysis of Simulation Results**
  - Model Verification
  - Model Validation
  - Transient Analysis
  - Terminating Simulations
  - Stopping Criteria
- Random Number Generation

Analysis of Simulation Results

- Issues to be resolved before simulation results can be used:
  - **Model verification:**
    - **Is the model implemented correctly?**
  - Model validation:
    - Is the model representative of the real system?
  - Transient removal:
    - How many initial observations to discard to ensure model has reached steady-state?
  - Stopping criteria:
    - How long to run the simulation?

Model Verification Techniques

- Techniques for debugging large programs
  - Code design:
    - Modularity
    - Top-down design
    - Anti-bugging
    - Structured walk-through
  - Test design:
    - Deterministic models/inputs
    - Simplified cases
    - Degeneracy tests (corner cases)
    - Consistency tests
    - Seed independence
  - Output design:
    - Traces (graphic form)
    - Continuity test

Model Validation Techniques

- Need to ensure that assumptions used in developing model are valid
- Validation techniques depend on the assumptions and the system
  - Not generally applicable across all systems
- Consist of validating three key aspects of the model
  - Assumptions
  - Input parameter values and distributions
  - Output values and conclusions
- Three possible sources of valid information (for comparison):
  - Expert intuition
  - Real system measurements
  - Theoretical results

Analysis of Simulation Results

- Issues to be resolved before simulation results can be used:
  - Model verification:
    - Is the model implemented correctly?
  - Model validation:
    - Is the model representative of the real system?
  - **Transient removal:**
    - **How many initial observations to discard to ensure model has reached steady-state?**
  - Stopping criteria:
    - How long to run the simulation?

Transient Removal

- In most simulations, only steady-state performance is of interest
  - Initial transient state of simulation should not be included in metric computations
- Main challenge:
  - Not possible to define exactly what constitutes the transient state
- Hence, all methods of transient removal are heuristic:
  - Long runs
  - Proper initialization
  - Truncation
  - Initial data deletion
  - Moving average of independent replications
  - Batch means

Long Runs

- Use long simulation runs:
  - Ensure that the presence of initial transients will not affect the result significantly
- ☹:
  - Wastes computation, storage, and processing resources
  - Difficult to ensure that the length of run chosen is long enough

Proper Initialization

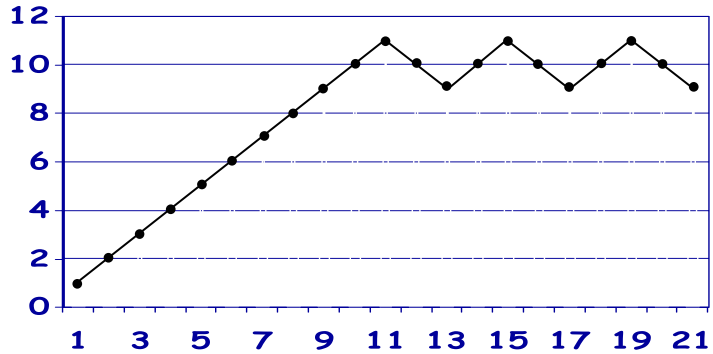
- Start the simulation in a state close to the expected steady state
  - Results in a reduction of length of transient period
- Example:
  - Start a CPU simulation with some jobs in the queue (not empty queue)
  - The number of initial jobs may be determined from previous simulations or by simple analysis

Truncation

- Assumption:
  - Variability during steady-state is less than during the transient state
- One measure of variability: range of observations
  - In a trajectory plotting successive observations, range of observations can often be seen to stabilize as simulation enters steady-state
- Method:
  - Given a sample {x1, x2, …, xn}
  - Ignore the first L observations and calculate the minimum and maximum of remaining (n-L) observations
  - Repeat above step for L=1, 2, …, n-1
    - Until (L+1)-th observation is neither the minimum nor maximum of the remaining observations.
  - Final value of L gives the length of transient state

Truncation Example

- Observations: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 10, 9, 10, 11, 10, 9, 10, 11, 10, 9



- At L=9, range of remaining sequence is: (9,11)
  - (L+1)-th observation is: 10
- Length of transient interval is, therefore, 9

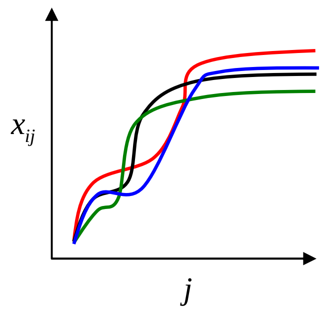
Initial Data Deletion

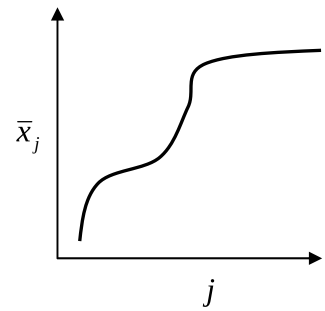
- Study change in overall average as some of the initial observations are deleted from the sample
- Assumption:
  - During steady-state, the average does not change much as some observations are deleted
- However, randomness does cause average to change slightly even in steady-state
  - First average across several repetitions
  - Watch for knee in the average-vs-observation plot

Initial Data Deletion: Methodology

- Suppose m replications of size n each
  - Let $\mathbf{x}_{ij}$ denote the j-th observation in the i-th replication
- Get a mean trajectory by averaging, for each j, across replications

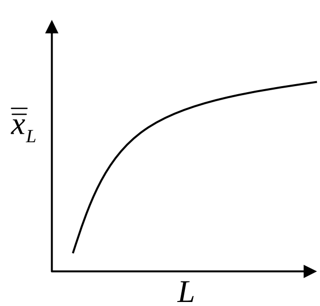
$$\bar{x}_j = \frac{1}{m}\sum_{i=1}^{m} x_{ij}$$

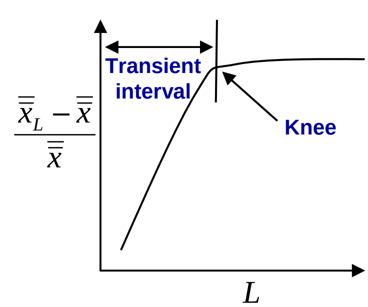




Initial Data Deletion: Methodology

- Get the overall mean of the **mean trajectory**: $\bar{\bar{x}} = \frac{1}{n}\sum_{j=1}^{n} \bar{x}_j$
- Get an overall mean of last (n-L) values of the **mean trajectory**: $\bar{\bar{x}}_L = \frac{1}{n-L}\sum_{j=L+1}^{n} \bar{x}_j$
- Compute the relative change in the overall mean: $\frac{\bar{\bar{x}}_L - \bar{\bar{x}}}{\bar{\bar{x}}}$
- Repeat for L= 1 to n-1
  - Knee of plot of (relative-change vs. L) gives length of the transient interval





Moving Average of Independent Replications

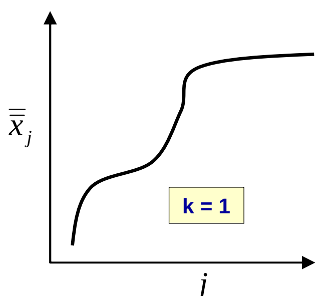
- Get a mean trajectory by averaging across replications:

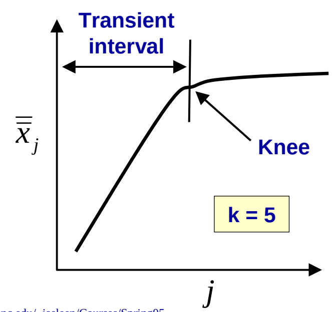
$$\bar{x}_j = \frac{1}{m}\sum_{i=1}^{m} x_{ij}, \quad j = 1,2,\ldots,n$$

- Plot trajectory of the moving average of successive (2k+1) values

$$\bar{\bar{x}}_j = \frac{1}{2k+1}\sum_{l=-k}^{k} \bar{x}_{j+l}, \quad j = k+1, k+2,\ldots,n-k$$

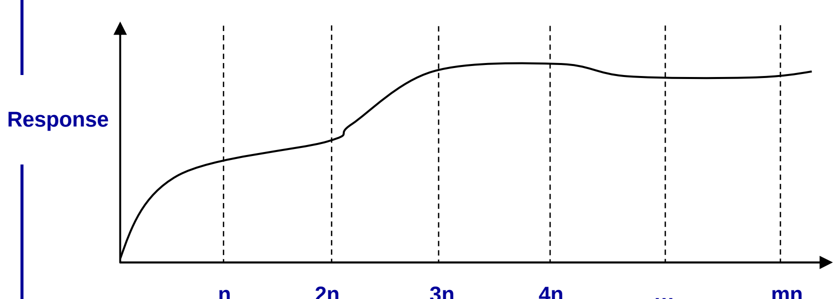
- Repeat with k = 1, 2, 3, … until the plot is sufficiently smooth
  - Knee of plot gives the length of transient phase





Batch Means

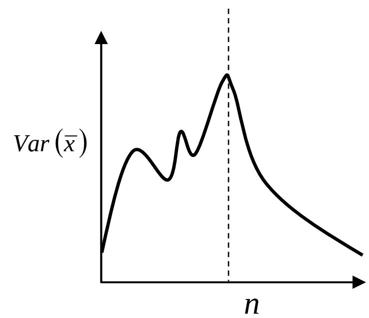
- Setup:
  - Run a very long simulation
  - Divide it up into several parts (**batches**) of equal duration



- Approach:
  - Study variance of batch means as a function of batch size

Batch Means: Methodology

- N observations can be divided into m batches of size n each
  - Let x_{ij} be the j-th observation in the i-th batch
- For each batch, compute a batch mean: $\bar{x}_i = \frac{1}{n}\sum_{j=1}^{n} x_{ij}, \quad i = 1,2,\ldots,m$
- Compute the overall mean: $\bar{\bar{x}} = \frac{1}{m}\sum_{i=1}^{m} \bar{x}_i$
- Compute the variance of the batch means: $Var(\bar{x}) = \frac{1}{m-1}\sum_{i=1}^{m} \left(\bar{x}_i - \bar{\bar{x}}\right)^2$
- Repeat for n=2, 3, 4, …
- Plot the variance as a function of n
  - Length of transient interval is the value of n at which the variance definitely starts decreasing



Batch Means: Rationale

- Rationale:
  - Suppose length of transient period is T
  - If n is much less than T
    - Initial batches bring the overall mean toward the initial batch means and the variance is small
  - As batch size increases, variance increases
  - At n > T, only first batch mean is different; others are roughly equal
    - Results in the decrease of variance

Other Issues Related to Steady-state

- Final conditions:
  - System state at the end of simulation may not be typical of the steady state
  - Exclude final portion of response from the steady-state computations
  - Methods similar to determining initial transient period

- Handling entities left at the end of the simulation:
  - Mean service time
    - Include only those jobs that have completed service
  - Mean waiting time:
    - Include only those jobs that have started execution
  - Mean queue length
    - Use time averages, instead of event averages

- Terminating simulations
  - Study system in transient state

Analysis of Simulation Results

- Issues to be resolved before simulation results can be used:
  - Model verification:
    - Is the model implemented correctly?
  - Model validation:
    - Is the model representative of the real system?
  - Transient removal:
    - How many initial observations to discard to ensure model has reached steady-state?
  - **Stopping criteria:**
    - **How long to run the simulation?**

Stopping Criteria

- Important to set length of simulation correctly
  - If too short, results may be highly variable
  - If too long, computing resources and manpower may be unnecessarily wasted
- Goal:
  - Run simulation until confidence interval for the mean response metric narrows to a desired width
- Confidence interval for the sample mean: $\bar{x} \pm z_{1-\alpha/2}\sqrt{Var\left(\bar{x}\right)}$
  - $z_{1-\alpha/2}$: derived from the unit normal variate
    - Values are well-known and tabulated
  - How to compute: $Var\left(\bar{x}\right)$

Variance of the Mean Response Metric

- Variance of the sample mean of n independent observations can be easily obtained from the variance of the observations as:

$$Var\left(\bar{x}\right) = Var\left(x\right)/n$$

  - Formula valid only if the observations are **independent**
- However, observations in most simulations are not independent
  - eg, successive waiting times are highly correlated
  - Formula can not be used to estimate variance of mean waiting time
- For correlated observations
  - Variance of the mean may be several times larger than that obtained from above formula !

Ignoring correlation may lead to narrower confidence intervals and premature termination of simulation

Computing Variance of Mean (Correlated Values)

- Several methods available:
  - Independent replications
  - Batch means
  - Regeneration
- Basic Idea:
  - Find regions of independence within the sample, to generate a set of independent means

Independent Replications

- Basic idea:
  - Mean of independent replications are independent
    - Even if observations within a replication are correlated
- Method uses m replications of size $(n + n_0)$ each
  - n_0: length of initial transient phase
  - First n_0 observations of each replication are discarded
- Suppose m replications of size n each
  - Let $\mathbf{x}_{ij}$ denote the j-th observation in the i-th replication

Independent Replications: Methodology

- For each replication, compute a mean: $\bar{x}_i = \frac{1}{n} \sum_{j=n_0+1}^{n_0+n} x_{ij}, \quad i = 1,2,\ldots,m$
- Compute the overall mean: $\bar{\bar{x}} = \frac{1}{m} \sum_{i=1}^{m} \bar{x}_i$
- Compute the variance of replicate means: $Var(\bar{x}) = \frac{1}{m-1} \sum_{i=1}^{m} (\bar{x}_i - \bar{\bar{x}})^2$
- Confidence interval for the mean response is: $\bar{\bar{x}} \pm z_{1-\alpha/2} \sqrt{Var(\bar{x})}$

Selecting m and n

- Method of independent replications requires discarding $m*n_0$ initial observations
  - Larger the m, larger is the wastage of data
- Confidence interval width is inversely proportional to: $\sqrt{mn}$
  - Narrower confidence interval can be obtained by increasing m or n
- Hence, it is recommended that:
  - Number of replications, m, should be kept fairly small (~10)
  - Length of replications, n, should be increased to obtain the desired confidence

Batch Means

- Setup:
  - Run a long simulation
  - Discard the initial transient interval
  - Divide the remaining observations into several batches
- Basic Idea:
  - If batch sizes are larger, means of batches are independent
- Given a long run of $N + n_0$ observations,
  - Where n_0 belong to initial transient phase and are discarded
  - **N** observations are divided into **m** batches of **n** observations each

Batch Means: Methodology

- For each batch, compute a mean: $\bar{x}_i = \frac{1}{n}\sum_{j=1}^{n} x_{ij}, \quad i = 1,2,\ldots,m$
- Compute the overall mean: $\bar{\bar{x}} = \frac{1}{m}\sum_{i=1}^{m} \bar{x}_i$
- Compute the variance of batch means: $Var(\bar{x}) = \frac{1}{m-1}\sum_{i=1}^{m} (\bar{x}_i - \bar{\bar{x}})^2$
- Confidence interval for the mean response is: $\bar{\bar{x}} \pm z_{1-\alpha/2}\sqrt{Var(\bar{x})}$

Selecting m and n

- Only n_0 initial observations are discarded
  - Independent of m or n
  - Wastage smaller than method of independent replications
- Confidence interval width is inversely proportional to: $\sqrt{mn}$
  - Narrower confidence interval can be obtained by increasing m or n
- Selecting n:
  - Batch size, n, must be large so that the batch means have little correlation
  - One idea: compute the auto-covariance of successive batch means:

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

  - Increase **n** till auto-covariance is small relative to variance of means

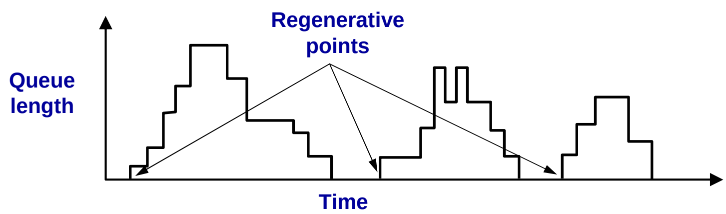
Example: Selecting Batch Size

Batch Size	Auto-covariance	Variance
1	-0.18792	1.79989
2	0.02643	0.81173
4	0.11024	0.42003
8	0.08979	0.26437
16	0.04001	0.17650
32	0.01108	0.10833
64	0.00010	0.06066
128	-0.00378	0.02992
256	0.00027	0.01133
512	0.00069	0.00503
1024	0.00078	0.00202

- At n=64, auto-covariance is less than 1% of original sample variance

Method of Regeneration

- Consider a single CPU scheduling simulation
  - Starts with an empty job queue



- System often returns to the initial state of empty job queue
  - On returning to this state, trajectory does not depend on past history
  - Regeneration Point
- Regeneration cycle:
  - Duration between two successive regeneration points

Regenerative Systems

- Mean metrics in different cycles are not correlated
  - Samples collected in different regenerative cycles are independent
- Variance computation is more complex than previous methods
  - Regeneration cycles are of different lengths
  - Overall mean response is not the average of the mean responses of individual cycles
    - If a regenerative simulation has **m** cycles of sizes: n_1, n_2, …n_m,

$$\bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}, \quad \bar{\bar{x}} \neq \frac{1}{m}\sum_{i=1}^{m} \bar{x}_i$$

Regenerative Systems: Confidence Intervals

- Compute cycle sums: $y_i = \sum_{j=1}^{n_i} x_{ij}, \quad i = 1,2,\ldots,m$
- Compute the overall mean: $\overline{\overline{x}} = \sum_{i=1}^{m} y_i \Big/ \sum_{i=1}^{m} n_i$
- Calculate difference between expected and observed cycle sums:
  $w_i = y_i - n_i \overline{\overline{x}}, \quad i = 1,2,\ldots,m$
- Calculate the variance of the difference: $Var(w) = s_w^2 = \frac{1}{m-1}\sum_{i=1}^{m}(w_i)^2$
- Compute the mean cycle length: $\overline{n} = \frac{1}{m}\sum_{i=1}^{m} n_i$
- Confidence interval for the mean response is: $\overline{\overline{x}} \pm z_{1-\alpha/2} \frac{s_w}{\overline{n}\sqrt{m}}$

Regenerative Systems Methodology

- ☺:
  - Does not require removing transient observations (no wastage of data)
- ☹:
  - Not all systems are regenerative
    - A system with two queues will need both queues to be empty
  - Cycle lengths are unpredictable
    - Not possible to plan the simulation time beforehand
  - Finding the regeneration point is not trivial
    - Requires checking after every event
  - Mean and variance estimators are biased
    - Expected values from a random sampling are not equal to the quantity being estimated

Simulations: Outline

- Simulation Basics
  - Terminology
  - Types of simulations
  - Implementing a simulator
- Analysis of Simulation Results
  - Model Verification
  - Model Validation
  - Transient Analysis
  - Terminating Simulations
  - Stopping Criteria
- **Random Number Generation**