Lecture 9: Output Data Analysis

Holger Füßler
Part A: Output Data Analysis for a Single System

Holger Füßler
Structure – Part A

» Part A.I: Problem statement ‘output analysis’
» Part A.II: Some probability theory and statistics
» Part A.III: Types of discrete event simulations
» Part A.IV: Credibility of simulation studies
A.I Problem statement

» By performing a simulation study one ‘observes’ a system.

» As output one gets a collection of data (statistics, traces).

» What ‘features’ can be inferred about the system?

“In many simulation studies a great deal of time and money is spent on model development and ‘programming,’ but little effort is made to analyze the simulation output data appropriately.”

“… a simulation is a computer-based statistical sampling experiment.”

[LK2000, Chap. 9]
A.I Problem statement cont‘d

» Do not get ,exact‘ answers

» Two different runs of the same model: different numerical results
A.1 Example 1: results for M/M/1 queue

Varying seeds (NetSim lab 3, change argument of lcgrand):

<table>
<thead>
<tr>
<th>Replication</th>
<th>Average delay</th>
<th>Average number in queue</th>
<th>Server utilization</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.606</td>
<td>0.060</td>
<td>0.192</td>
</tr>
<tr>
<td>2</td>
<td>0.554</td>
<td>0.057</td>
<td>0.206</td>
</tr>
<tr>
<td>3</td>
<td>0.586</td>
<td>0.057</td>
<td>0.200</td>
</tr>
<tr>
<td>4</td>
<td>0.452</td>
<td>0.046</td>
<td>0.197</td>
</tr>
<tr>
<td>5</td>
<td>0.490</td>
<td>0.050</td>
<td>0.199</td>
</tr>
</tbody>
</table>
# A.I Example 1: results for M/M/1 queue cont‘d

» Varying number of packets (NetSim lab 3, change parameter in mm1.in)

<table>
<thead>
<tr>
<th>Number of packets</th>
<th>Average delay in queue</th>
<th>Average number in queue</th>
<th>Server utilization</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>0.606</td>
<td>0.060</td>
<td>0.192</td>
</tr>
<tr>
<td>2000</td>
<td>0.554</td>
<td>0.057</td>
<td>0.206</td>
</tr>
<tr>
<td>3000</td>
<td>0.586</td>
<td>0.057</td>
<td>0.200</td>
</tr>
<tr>
<td>4000</td>
<td>0.452</td>
<td>0.046</td>
<td>0.197</td>
</tr>
<tr>
<td>5000</td>
<td>0.490</td>
<td>0.050</td>
<td>0.199</td>
</tr>
</tbody>
</table>

» What is the 'true' value?

» How much does the obtained result differ from the 'true' value?
A.1 Example 2: TCP average achievable bandwidth

» How long is the warm-up period?

» Assume that we have random bit/link errors: what is the achievable bandwidth?
A.1 Example 2: TCP achievable bandwidth

Conversion to meaningful results?
A.1 General set-up

\[ Y_{11} \quad \ldots \quad Y_{1i} \quad \ldots \quad Y_{1m} \]
\[ Y_{21} \quad \ldots \quad Y_{2i} \quad \ldots \quad Y_{2m} \]
\[ \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \]
\[ Y_{n1} \quad \ldots \quad Y_{ni} \quad \ldots \quad Y_{nm} \]

\text{Challenge: } Choose \ n, \ m \ \text{and warm-up period appropriately!}
A.II Estimates of means and variances

Suppose that $X_1, X_2, \ldots, X_n$ are IID random variables with finite mean $\mu$ and finite variance $\sigma^2$.

How can we get an estimate for $\mu$ and $\sigma$?

Unbiased estimator for $\mu$: sample mean

$$\overline{X}(n) = \frac{\sum_{i=1}^{n} X_i}{n}$$

Unbiased estimator for $\sigma^2$: sample variance

$$S^2(n) = \frac{\sum_{i=1}^{n} [X_i - \overline{X}(n)]^2}{n - 1}$$
A.II Estimates of means and variances (Calculation)

let $X_j^\Sigma$ be the sum of all values until the $j$-th element and $X_j^{\Sigma^2}$ the sum of the respective squares, i.e.

$$X_j^\Sigma = \sum_{i=1}^{j} X_i \quad X_j^{\Sigma^2} = \sum_{i=1}^{j} X_i^2$$

then, the estimators can be calculated as

$$\bar{X}(n) = \frac{X_n^\Sigma}{n}$$

$$S^2(n) = \frac{1}{n - 1} \left( X_n^{\Sigma^2} - 2 \cdot (\bar{X}(n)) \cdot X_n^\Sigma + n \cdot (\bar{X}(n))^2 \right)$$

$\Rightarrow$ we only need the accumulated sum / squared sum of each variable
A.II Confidence intervals for the mean

» Again, assume \( X_1, X_2, \ldots, X_n \) are IID random variables with finite mean \( \mu \) and finite variance \( \sigma^2 \) greater 0.

» Central limit theorem states: \( \bar{X}(n) \) is approximately distributed as a normal random variable with mean \( \mu \) and variance \( \sigma^2/n \).

» For sufficiently large \( n \), an approximate 100(1-\( \alpha \)) percent confidence interval for \( \mu \) is given by

\[
\bar{X}(n) \pm z_{1-\alpha/2} \sqrt{S^2(n)/n}
\]

» Interpretation of ‘confidence interval’: “… in 100(1-\( \alpha \)) percent of all cases the true parameter \( \mu \) is within the interval.

» Why ‘approximate’?
  - It is only asymptotically correct

Warning: Only „good“ for sample size of approx. > 50
A.II Confidence intervals for the mean: illustration

Assumption: N(0,1)
### A.II Example: M/M/1 queue

<table>
<thead>
<tr>
<th>Replication</th>
<th>Average delay</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.606</td>
</tr>
<tr>
<td>2</td>
<td>0.554</td>
</tr>
<tr>
<td>3</td>
<td>0.586</td>
</tr>
<tr>
<td>4</td>
<td>0.452</td>
</tr>
<tr>
<td>5</td>
<td>0.490</td>
</tr>
<tr>
<td>6</td>
<td>0.548</td>
</tr>
<tr>
<td>7</td>
<td>0.519</td>
</tr>
<tr>
<td>8</td>
<td>0.498</td>
</tr>
<tr>
<td>9</td>
<td>0.366</td>
</tr>
<tr>
<td>10</td>
<td>0.364</td>
</tr>
<tr>
<td><strong>Average (.3digits)</strong></td>
<td><strong>0.498</strong></td>
</tr>
</tbody>
</table>

- Experiments for 1000 packets each
- \( S^2(10) = 0.007 \)
- 95\% confidence interval:
  - \( z \approx 2.0 \)
  - \( 0.498 \pm 0.014 \)
A.II Selecting the sample size

Result so far (under a lot of assumptions): for sufficiently large $n$, an approximate $100(1-\alpha)$ percent confidence interval for $\mu$ is given by

$$\bar{X}(n) \pm z_{1-\alpha/2} \sqrt{\frac{S^2(n)}{n}}$$

How do we have to select the number of samples $n$?

Let us bound this by an absolute value $v$

Thus, given $S^2(n)$ (or $\sigma^2$ if it is known) and value $v$ and $z_{1-\alpha/2}$, one can solve for number of samples $n$. 
A.II Example: calculating required sample sizes

» Let \( Y_i, i=1, 2, \ldots \), be IID Bernoulli random variables with parameter \( p \).

» What is the sample size necessary to estimate \( p \) within 0.05 with probability .95?

» Assume no information is given w.r.t. the variance.

» Max. variance: 0.025

» \( 0.025 \cdot (2/0.05)^2 = n \)

» \( n=400 \)
A.II Experiment: Estimated coverages

» Coverage: proportion of confidence intervals that contain the 'true' parameter $\mu$.

» Should be $1 - \alpha$ for 'n sufficiently large'

» Can be checked for known distributions.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Skewness</th>
<th>M=5</th>
<th>M=10</th>
<th>M=20</th>
<th>M=40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>0.00</td>
<td>0.910</td>
<td>0.902</td>
<td>0.898</td>
<td>0.900</td>
</tr>
<tr>
<td>Exponential</td>
<td>2.00</td>
<td>0.854</td>
<td>0.878</td>
<td>0.870</td>
<td>0.890</td>
</tr>
<tr>
<td>Chi square</td>
<td>2.83</td>
<td>0.810</td>
<td>0.830</td>
<td>0.848</td>
<td>0.890</td>
</tr>
<tr>
<td>Lognormal</td>
<td>6.18</td>
<td>0.758</td>
<td>0.768</td>
<td>0.842</td>
<td>0.852</td>
</tr>
<tr>
<td>Hyperexp.</td>
<td>6.43</td>
<td>0.584</td>
<td>0.586</td>
<td>0.682</td>
<td>0.774</td>
</tr>
</tbody>
</table>

Estimated coverages for 90 percent confidence intervals based on 500 independent experiments for each of the sample sizes [Source: Law/Kelton]
A.III Types of simulations w.r.t. output analysis

Terminating simulation

Nonterminating simulation

Steady-state parameter

Steady-state cycle parameter

Terminating: Parameters to be estimated are defined relative to specific initial and stopping conditions that are part of the model.

Nonterminating: There is no natural and realistic event that terminates the model. Interested in “long-run” behavior characteristic of “normal” operation. If the performance measure of interest is a characteristic of a steady-state distribution of the process, it is a steady-state parameter of the model.

Not all nonterminating systems are steady-state: there could be a periodic “cycle” in the long run, giving rise to steady-state cycle parameters.
A.III Types of simulations w.r.t. output analysis

» Terminating simulations: „9 to 5 scenarios“

» Example: M/M/1 queue
  – Initial condition: empty queue
  – Terminating condition: time elapsed

» Statistics for terminating simulations: see Part II of this lecture

» Challenge: steady-state simulations
  – How to get rid of impact of initial condition?
  – When to stop simulation?
A.IV „Crisis of credibility“


(c) IEEE/ACM Transactions on Networking
A.IV Recommendations (Pawlikowski et al.)

» Reported simulation experiments should be repeatable
  – Give information about
    • The PRNG(s) used during the simulation
    • The type of simulation
    • The method of analysis of simulation output data
    • The final statistical errors associated with the results

Indicate confidence interval (CI) for specified confidence level (CL)
A.IV Comments for 'best practices' 

» Independence or covariance-stationarity rarely encountered in practice 😊

» But: if the number of replications, samples etc. is too low even under the assumptions of independence or covariance-stationarity, something is probably flawed …
   – We need mathematical results to check

» In reality, also time and space constraints can severely impact achievable confidence intervals
   – But this should be specified

» Trace and plot as many variables as possible to cross-check correctness 😊
A.IV Wrap-up

» Any stochastic computer simulation (using RNGs/(PRNGs) has to be regarded as a (simulated) statistical experiment.

» Statistics background:
  – estimating means and variances
  – confidence levels and intervals
  – hypothesis testing

» Transient and steady-state behavior

» Terminating, steady-state and cyclic steady state simulations

» The issue of credibility
References – Part A

  – Chapters 4 and 9
Part B: Comparing Different Configurations
Where we are …

Why network simulations?

» Educational use
  – See protocol in action
  – Does it work as intended?

» Get some quantitative results for a single configuration
  – E.g., how long does it take to find a route?

» Compare different configuration and decide which one is ‘better’
  – Trade-offs
Lecture overview – Part B

» Part B.I: Scope of this Lecture - Motivation
» Part B.II: Comparison of different configurations
» Part B.III: Variance reduction techniques
Which one is ‘better’?
- Answer depends on metric(s)
- Answer depends on scenario
  - Mobility pattern
  - Communication pattern
  - Caching strategies
  - Flooding strategies
  - ...

Metrics:
- Packet delivery ratio
- Route acquisition time
- End-to-end delay
- Overhead costs
B.II Comparison of two different configurations

» Let’s go back to some simple scenario

» Compare M/M/1 queue (service time: exponential with mean 0.9) with M/M/2 queue (service time: exponential with mean 1.8 each)
### B.II Motivation by example 2

<table>
<thead>
<tr>
<th>Experiment</th>
<th>$d_z(100)$</th>
<th>$d_k(100)$</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.80</td>
<td>4.60</td>
<td>Zippy (wrong)</td>
</tr>
<tr>
<td>2</td>
<td>3.17</td>
<td>8.37</td>
<td>Zippy (wrong)</td>
</tr>
<tr>
<td>3</td>
<td>3.96</td>
<td>4.18</td>
<td>Zippy (wrong)</td>
</tr>
<tr>
<td>4</td>
<td>1.91</td>
<td>5.77</td>
<td>Zippy (wrong)</td>
</tr>
<tr>
<td>5</td>
<td>1.71</td>
<td>2.23</td>
<td>Zippy (wrong)</td>
</tr>
<tr>
<td>6</td>
<td>6.16</td>
<td>4.72</td>
<td>Klunky (right)</td>
</tr>
<tr>
<td>7</td>
<td>5.67</td>
<td>1.39</td>
<td>Klunky (right)</td>
</tr>
<tr>
<td>98</td>
<td></td>
<td></td>
<td>Zippy (wrong)</td>
</tr>
<tr>
<td>99</td>
<td></td>
<td></td>
<td>Klunky (right)</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
<td>Klunky (right)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n$</th>
<th>$P$(wrong answer)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.52</td>
</tr>
<tr>
<td>5</td>
<td>0.43</td>
</tr>
<tr>
<td>10</td>
<td>0.38</td>
</tr>
<tr>
<td>20</td>
<td>0.34</td>
</tr>
</tbody>
</table>
B.II Motivation by example 3

![Simulation of Computer Networks](image)

Simulated Expected
2 Klunkies ·····
1 Zippy ·····

Average delay in queue

\( n = 20 \)
\( n = 10 \)
\( n = 5 \)
\( n = 1 \)
B.II Confidence interval for the difference between two systems

» Two alternative simulated systems \((i = 1, 2)\), \(\mu_i = \) expected performance measure from system \(i\)

» Take “sample” of \(n_i\) observations (replications) from system \(i\)

» \(X_{ij} = \) observation \(j\) from system \(i\)

» Want: confidence interval on \(z = \mu_1 - \mu_2\)

» If interval misses \(0\), conclude there is a *statistical* difference between the systems

» Is the difference *practically* significant? Must use judgment in context.
B.II Paired confidence interval

» Assume \( n_1 = n_2 \) (=n, say)

» For a fixed \( j \), \( X_{1j} \) and \( X_{2j} \) need not be independent
  – Important for variance reduction techniques (next part)

» Let \( Z_j = X_{1j} - X_{2j} \)

» Problem reduced to ‘single system problem’

» Find confidence interval for \( E[Z_j] \)

» Previous example (10 runs):

\[
\bar{Z}(10) = 0.376 \\
S^2(10)/90 \approx 1.25 \\
\text{Confidence interval for CL 95\%: } 0.376 \pm 2.24
B.II Issues not covered in this part

» Other comparison methods

» Comparing more than two systems

» Ranking and selection
B.III Variance reduction techniques

» Main drawback of using simulation to study stochastic models:
   Results are uncertain — have variance associated with them

» Would like to have as little variance as possible — more precise results

» One sure way to decrease the variance:
   Run it some more (longer runs, additional replications)

» Sometimes can manipulate simulation to reduce the variance of the output at little or no additional cost — not just by running it some more

» Another way of looking at it — try to achieve a desired level of precision (e.g., confidence-interval smallness) with less simulating — Variance-reduction technique (VRT)
B.III Common random numbers (CRN)

- When comparing two or more alternative system configurations

- Basic idea: compare alternative configurations ‘under similar experimental conditions’ – use random numbers ‘for same purpose’
  - Often used ‘unconsciously’
  - E.g. use same movement and communication pattern when comparing two ad-hoc routing protocols

- Example of ‘what can go wrong’:

<table>
<thead>
<tr>
<th>$k$</th>
<th>$U_k$</th>
<th>Usage in M/M/1</th>
<th>Usage in M/M/2</th>
<th>Agree?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.401</td>
<td>A</td>
<td>A</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>0.614</td>
<td>A</td>
<td>A</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>0.434</td>
<td>S</td>
<td>S</td>
<td>Yes</td>
</tr>
<tr>
<td>4</td>
<td>0.383</td>
<td>A</td>
<td>A</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>0.506</td>
<td>S</td>
<td>S</td>
<td>Yes</td>
</tr>
<tr>
<td>6</td>
<td>0.709</td>
<td>A</td>
<td>A</td>
<td>Yes</td>
</tr>
<tr>
<td>7</td>
<td>0.185</td>
<td>S</td>
<td>S</td>
<td>Yes</td>
</tr>
<tr>
<td>8</td>
<td>0.834</td>
<td>A</td>
<td>A</td>
<td>Yes</td>
</tr>
<tr>
<td>9</td>
<td>0.646</td>
<td>A</td>
<td>S</td>
<td>No</td>
</tr>
<tr>
<td>10</td>
<td>0.376</td>
<td>A</td>
<td>A</td>
<td>Yes</td>
</tr>
</tbody>
</table>
B.III Mathematical basis for CRN

We have two alternatives, where $X_{1j}$ and $X_{2j}$ are the observations from the first and second configuration on the jth independent replication.

Again, let $Z_j = X_{1j} - X_{2j}$.

$$Var[\bar{Z}(n)] = \frac{Var(Z_j)}{n} = \frac{Var(X_{1j}) + Var(X_{2j}) - 2Cov(X_{1j}, X_{2j})}{n}$$

When we can induce some positive correlation, we can make … smaller.
B.III Applicability of CRN
B.III CRN – Synchronization techniques

» Use of dedicated random number streams

» Use of inverse transform
  – But: inverse transform not always the most efficient choice …

» Compute random numbers in advance
  – Costs some memory

» … or waste some random numbers
  – … to keep things synchronized
B.III CRN at work 1

Independent Sampling

$M/M/1 (X_{1j})$  
$M/M/2 (X_{2j})$

CRN (A & S)

$M/M/1 (X_{1j})$  
$M/M/2 (X_{2j})$

Replication ($j$)

Simulation of Computer Networks  
Holger Füßler - 42  
B.III CRN at work 2

\[ X_{ij}, Z_j \]

- Independent \( \triangle \)
- CRN (A & S) \( \blacktriangle \)

\( n = 20 \)
\( n = 10 \)
\( n = 5 \)
\( n = 1 \)
B.IV Application to network simulations

» When comparing two alternatives
  – Use ‘same’ topology
    • E.g. preprocessed movement pattern, same radio transmission range
    • E.g. same (preprocessed) link error patterns
  – Use ‘same’ communication pattern

» What else is ‘random’ and can affect results?
Wrap-up Part B

» Today’s focus: comparison of two alternative configurations

» Problem reduced to finding a confidence interval of a ‘single’ system

» Confidence intervals: computation for specified precision

» Precision corresponds to variance (of sample mean): variance reduction techniques needed
  – Common random numbers
References – Part B

» Law, Kelton: Chapters 10 and 11