2WB05 Simulation
Lecture 7: Output analysis

Marko Boon
http://www.win.tue.nl/courses/2WB05
Outline

Output analysis of a simulation
- Confidence intervals
- Warm-up interval
- Common random numbers
Confidence intervals

Let $X_1, X_2, \ldots, X_n$ be independent realizations of a random variable $X$ with unknown mean $\mu$ and unknown variance $\sigma^2$.

Sample mean

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

Sample variance

$$S^2(n) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}(n))^2$$

Clearly $\bar{X}(n)$ is an estimator for the unknown mean $\mu$.

How can we construct a confidence interval for $\mu$?
Confidence intervals

Central limit theorem states that for large $n$

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma \sqrt{n}}$$

is approximately a standard normal random variable, and this remains valid if $\sigma$ is replaced by $S(n)$. Hence, let $z_\beta = \Phi^{-1}(\beta)$ (e.g., $z_{1-0.025} = 1.96$), then

$$P\left(-z_{1-\alpha/2} \leq \frac{\sum_{i=1}^{n} X_i - n\mu}{S(n) \sqrt{n}} \leq z_{1-\alpha/2}\right) \approx 1 - \alpha$$

or equivalently

$$P\left(\bar{X}(n) - z_{1-\alpha/2} \frac{S(n)}{\sqrt{n}} \leq \mu \leq \bar{X}(n) + z_{1-\alpha/2} \frac{S(n)}{\sqrt{n}}\right) \approx 1 - \alpha$$
Conclusion

An approximate $100(1 - \alpha)$% confidence interval for the unknown mean $\mu$ is given by

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

As a consequence, to obtain one extra digit of the parameter $\mu$, the required simulation time increases with approximately a factor 100.
100 confidence intervals for the mean of uniform random variable on \((-1, 1)\); each interval is based on 100 observations.
Remark:
If the observations $X_i$ are normally distributed, then

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{S(n)\sqrt{n}}$$

has for all $n$ a Student’s $t$ distribution with $n - 1$ degrees of freedom; so an exact confidence interval can be obtained by replacing $z_{1-\alpha/2}$ by the corresponding quantile of the $t$ distribution with $n - 1$ degrees of freedom.
Confidence intervals

Remark:
Recursive computation of the sample mean and variance of the realizations $X_1, \ldots, X_n$ of a random variable $X$:

$$
\bar{X}(n) = \frac{n - 1}{n} \bar{X}(n - 1) + \frac{1}{n} X_n
$$

and

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

for $n = 2, 3, \ldots$, where

$$
\bar{X}(1) = X_1, \quad S^2(1) = 0.
$$
Problem of the initialization effect

We are interested in the long-term behaviour of the system and maybe the choice of the initial state of the simulation will influence the quality of our estimate. One way of dealing with this problem is to choose $N$ very large and to neglect this initialization effect. However, a better way is to throw away in each run the first $k$ observations.
Let $W_1, W_2, \ldots, W_N$ be realizations of waiting times in a single run, and suppose we want to estimate the steady-state mean waiting time $E(W)$, defined as

$$E(W) = \lim_{j \to \infty} E(W_j)$$

by the sample mean

$$\bar{W}_N = \frac{1}{N} \sum_{j=1}^{N} W_j$$
In this estimate there are two types of errors:

- **Systematic error, or bias**
  This means that
  \[ E(\bar{W}_N) \neq E(W), \]
  due to the influence of the initial conditions, which may not be “representative” for steady-state behavior;

- **Sampling (or random) error**
  The estimator \( \bar{W}_N \) is of course a random variable.
To reduce the systematic error, we delete the initial observations, say $W_1, \ldots, W_k$, and use the remaining observations $W_{k+1}, \ldots, W_N$ to estimate $E(W)$ by the truncated sample mean

$$\bar{W}_{k,N} = \frac{1}{N-k} \sum_{j=k+1}^{N} W_j$$

Then one expects that $\bar{W}_{k,N}$ is less biased than $\bar{W}_N$, since the observations near the beginning of the simulation may not be representative for steady-state behavior; the parameter $k$ is called the warm-up interval.
Choosing the warm-up interval

We like to pick $k$ such that $E(\bar{W}_{k,N}) \approx E(W)$.

- If $k$ is too small, then $E(\bar{W}_{k,N})$ may be significantly different from $E(W)$;
- If $k$ is too large, then the variance of $\bar{W}_{k,N}$ (the sampling error) may be too large (its variance is proportional to $1/(N - k)$).
Graphical procedure

Our goal is to determine a value $k$ such that

$$E(W_j) \approx E(W)$$

for all $j > k$.

The presence of variability of the process $W_1, W_2, \ldots$ makes it hard to determine $k$ from a single run.

Therefore, the idea is to make $n$ independent replications (by using different random numbers) and employing the following steps:
1. Make \( n \) independent replications (or runs), each of length \( N \); let \( W_j^{(i)} \) denote the \( j \)-th waiting time in run \( i \).

2. Let

\[
\bar{W}_j = \frac{1}{n} \sum_{i=1}^{n} W_j^{(i)}
\]

The averaged process \( \bar{W}_1, \bar{W}_2, \ldots \) has means and variances

\[
E(\bar{W}_j) = E(W_j), \quad \text{var}(\bar{W}_j) = \frac{\text{var}(W_j)}{n}.
\]

So its mean behavior is the same as the original process, but it has a smaller \((1/n\)-th\) variance.

3. Plot \( \bar{W}_j \) and choose \( k \) such that beyond \( k \) the process \( \bar{W}_1, \bar{W}_2, \ldots \) appears to have converged.
Waiting times for the $M/M/1$ queue with $\lambda = 0.5$ and $\mu = 1$;
Averaged waiting times for the $M/M/1$ queue with $\lambda = 0.5$ and $\mu = 1$; number of replications is 5.
Averaged waiting times for the $M/M/1$ queue with $\lambda = 0.5$ and $\mu = 1$; number of replications is 10.
Warm-up interval

Averaged waiting times for the $M/M/1$ queue with $\lambda = 0.5$ and $\mu = 1$; number of replications is 100.
Warm-up interval

To smooth “high-frequency” oscillations in \( \tilde{W}_1, \tilde{W}_2, \ldots \) (but leave the trend) one may consider the moving average \( \tilde{W}_j(w) \) (where \( w \) is the window size) defined as:

\[
\tilde{W}_j(w) = \frac{1}{2w + 1} \sum_{i=-w}^{w} \tilde{W}_{j+i}
\]

for \( j = w + 1, w + 2, \ldots, N - w \), and

\[
\tilde{W}_j(w) = \frac{1}{2j - 1} \sum_{i=-(j-1)}^{j-1} \tilde{W}_{j+i}
\]

for \( j = 1, \ldots, w \).

The warm-up interval \( k \) can then be determined from the plot of \( \tilde{W}_j(w) \) for \( j = 1, \ldots, N - w \).
Moving average of averaged waiting times for the $M/M/1$ queue with $\lambda = 0.5$ and $\mu = 1$; number of replications is 10 and window size is 5.
Warm-up interval

Moving average of averaged waiting times for the $M/M/1$ queue with $\lambda = 0.5$ and $\mu = 1$; number of replications is 10 and window size is 10.
Moving average of averaged waiting times for the $M/M/1$ queue with $\lambda = 0.5$ and $\mu = 1$; number of replications is 100 and window size is 10.
Batch means

Instead of doing $n$ independent runs, we try to obtain $n$ independent observations by making a single long run and, after deleting the first $k$ observations, dividing this run into $n$ subruns.

The advantage is that we have to go through the warm-up period only once.
Let $W_1, W_2, \ldots, W_{nN}$ be the output of a single run, where we have already deleted the first $k$ observations and renumbered the remaining ones. Hence $W_1, W_2, \ldots, W_{nN}$ will be representative for the steady-state. We divide the observations into $n$ batches of length $N$. Thus, batch 1 consists of

$$W_1, W_2, \ldots, W_N;$$

batch 2 of

$$W_{N+1}, W_{N+2}, \ldots, W_{2N},$$

and so on. Let $\bar{W}_N^{(i)}$ be the sample (or batch) mean of the $N$ observations in batch $i$, so

$$\bar{W}_N^{(i)} = \frac{1}{N} \sum_{j=(i-1)N+1}^{iN} W_j$$
The $\bar{W}_N^{(i)}$’s play the same role as the ones in the independent replication method. Unfortunately, the $\bar{W}_N^{(i)}$’s will now be dependent.

But, under mild conditions, for large $N$ the $\bar{W}_N^{(i)}$’s will be approximately independent, each with the same mean $E(W)$.

Hence, for $N$ large enough, it is reasonable to treat the $\bar{W}_N^{(i)}$’s as i.i.d. random variables with mean $E(W)$; thus

$$\bar{W}_{n,N} \pm z_{1-\delta/2} \frac{S_{n,N}}{\sqrt{n}}$$

provides again a $100(1-\delta)\%$ confidence interval for $E(W)$, with $\bar{W}_{n,N}$ and $S_{n,N}^2$ again the sample mean and variance of the realizations $\bar{W}_N^{(1)}, \ldots, \bar{W}_N^{(n)}$. 
Common random numbers

If we compare the performance of two systems with random components it is in general better to evaluate both systems with the same realizations of the random components.

If $X$ and $Y$ are estimators for the performance of the two systems, then

$$\text{var}(X - Y) = \text{var}(X) + \text{var}(Y) - 2 \text{cov}(X, Y).$$

In general, use of common random numbers leads to positively correlated $X$ and $Y$:

$$\text{cov}(X, Y) > 0$$

Hence, the variance of $X - Y$ (and thus the corresponding confidence interval) will be smaller than in case $X$ and $Y$ are independent (when generated with independent random numbers).
Suppose that $N$ jobs have to be processed on $M$ identical machines. The processing times are independent and exponentially distributed with mean 1.

We want to compare the completion time of the last job, $C_{\text{max}}$, under two different strategies:

- Longest Processing Time First (LPTF)
- Shortest Processing Time First (SPTF)

**Remark:** Let $X_1, \ldots, X_N$ be independent exponentials with mean 1; denote the smallest one by $X_{(1)}$, the second smallest one by $X_{(2)}$, and so on.

Then for $i = 1, \ldots, N$

$$X_{(i)} \overset{d}{=} Y_N + \cdots + Y_{N-i+1}$$

where $Y_1, \ldots, Y_N$ are independent exponentials with $E(Y_i) = 1/i$ ($Y_i$ is the minimum of $i$ exponentials with mean 1).
Example: Job scheduling

Results for $M = 2$, $N = 10$ *without* using common jobs for the SPTF and LPTF strategy ($n$ is the number of experiments):

<table>
<thead>
<tr>
<th>$n$</th>
<th>$C_{\text{SPTF}}^\text{max} - C_{\text{LPTF}}^\text{max}$ mean</th>
<th>st.dev.</th>
<th>half width</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^3$</td>
<td>0.860</td>
<td>2.43</td>
<td>0.151</td>
<td></td>
</tr>
<tr>
<td>$10^4$</td>
<td>0.867</td>
<td>2.49</td>
<td>0.049</td>
<td></td>
</tr>
<tr>
<td>$10^5$</td>
<td>0.849</td>
<td>2.50</td>
<td>0.016</td>
<td></td>
</tr>
</tbody>
</table>

and the results using common jobs:

<table>
<thead>
<tr>
<th>$n$</th>
<th>$C_{\text{SPTF}}^\text{max} - C_{\text{LPTF}}^\text{max}$ mean</th>
<th>st.dev.</th>
<th>half width</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^3$</td>
<td>0.840</td>
<td>0.497</td>
<td>0.031</td>
<td></td>
</tr>
<tr>
<td>$10^4$</td>
<td>0.833</td>
<td>0.512</td>
<td>0.010</td>
<td></td>
</tr>
<tr>
<td>$10^5$</td>
<td>0.838</td>
<td>0.518</td>
<td>0.003</td>
<td></td>
</tr>
</tbody>
</table>

Hence, using common jobs reduces the (sample) standard deviation of $C_{\text{SPTF}}^\text{max} - C_{\text{LPTF}}^\text{max}$, and thus the width of the confidence interval for its mean with a factor 5!
Example: Job scheduling

Results for $M = 2$, $N = 10$ without using common jobs for the SPTF and LPTF strategy ($n$ is the number of experiments):

<table>
<thead>
<tr>
<th>$n$</th>
<th>$C_{\text{max}}^{\text{SPTF}} - C_{\text{max}}^{\text{LPTF}}$</th>
<th>mean</th>
<th>st.dev.</th>
<th>half width 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.860</td>
<td>2.43</td>
<td>0.151</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.867</td>
<td>2.49</td>
<td>0.049</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.849</td>
<td>2.50</td>
<td>0.016</td>
<td></td>
</tr>
</tbody>
</table>

and the results using common jobs:

<table>
<thead>
<tr>
<th>$n$</th>
<th>$C_{\text{max}}^{\text{SPTF}} - C_{\text{max}}^{\text{LPTF}}$</th>
<th>mean</th>
<th>st.dev.</th>
<th>half width 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.840</td>
<td>0.497</td>
<td>0.031</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.833</td>
<td>0.512</td>
<td>0.010</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.838</td>
<td>0.518</td>
<td>0.003</td>
<td></td>
</tr>
</tbody>
</table>

Hence, using common jobs reduces the (sample) standard deviation of $C_{\text{max}}^{\text{SPTF}} - C_{\text{max}}^{\text{LPTF}}$, and thus the width of the confidence interval for its mean with a factor $5!$

Conclusion: common random numbers is better!

Example: Job scheduling
We want to determine the reduction in the mean waiting time when we have an extra machine. To compare the two situations we want to use the same realizations of arrival times and processing times in the simulation.

Assume that processing times of jobs are generated when they enter production. Then, the order in which arrival and processing times are generated depends on the number of machines: synchronization problem.

Solution:
- Use separate random number streams for different random variables (arrival and processing times);
- Design the simulation model such that it guarantees that exactly the same realizations of random variables are generated.

In this problem, the second approach can be realized by assigning to each job a processing time immediately upon arrival.
Example: Production system

Results for $\lambda = 4, \mu = 1$ and the number of machines $M$ is 5, resp. 6. In each run $N = 10^5$ waiting times are generated;

<table>
<thead>
<tr>
<th>$i$</th>
<th>$\overline{W}_N^{(i)}$</th>
<th>$\Delta^{(i)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M = 5$</td>
<td>$M = 6$</td>
</tr>
<tr>
<td>1</td>
<td>0.607</td>
<td>0.150</td>
</tr>
<tr>
<td>2</td>
<td>0.545</td>
<td>0.138</td>
</tr>
<tr>
<td>3</td>
<td>0.527</td>
<td>0.139</td>
</tr>
<tr>
<td>4</td>
<td>0.526</td>
<td>0.135</td>
</tr>
<tr>
<td>5</td>
<td>0.595</td>
<td>0.157</td>
</tr>
<tr>
<td>6</td>
<td>0.569</td>
<td>0.144</td>
</tr>
<tr>
<td>7</td>
<td>0.587</td>
<td>0.150</td>
</tr>
<tr>
<td>8</td>
<td>0.577</td>
<td>0.149</td>
</tr>
<tr>
<td>9</td>
<td>0.553</td>
<td>0.145</td>
</tr>
<tr>
<td>10</td>
<td>0.554</td>
<td>0.138</td>
</tr>
</tbody>
</table>

The standard deviation of the $\Delta^{(i)}$ is equal to 0.022. If both systems use independent realizations of arrival and processing times the standard dev. of $\Delta^{(i)}$ is 0.029; so common random numbers yields a reduction of 25%.
Results for $\lambda = 4$, $\mu = 1$ and the number of machines $M$ is 5, resp. 6. In each run $N = 10^5$ waiting times are generated;

<table>
<thead>
<tr>
<th>$i$</th>
<th>$\bar{W}_N^{(i)}$</th>
<th>$\Delta^{(i)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M = 5$</td>
<td>$M = 6$</td>
<td></td>
</tr>
</tbody>
</table>

Blah blah blah ...
Merry Christmas!

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.554</td>
<td>0.138</td>
<td>0.416</td>
</tr>
</tbody>
</table>

The standard deviation of the $\Delta^{(i)}$ is equal to 0.022. If both systems use independent realizations of arrival and processing times the standard dev. of $\Delta^{(i)}$ is 0.029; so common random numbers yields a reduction of 25%.