Chapter –IV: Design of Simulation Experiments

Email:- begna19mrblgo@gmail.com

Design of Simulation Experiments

- Point of most simulation projects: running the model and trying to understand the results.
- Need to plan ahead how this will be done, to avoid time-consuming inefficiencies.
- A simulation study is an experiment that needs to be designed.
- Identify purpose of the project Just one system configuration?
- Still have issues like run length, number of runs, how to assign random numbers driving the simulation, interpretation of results Several "given" system configurations?
- Have same questions, as well as how to compare, select, or rank the alternatives.
- How do changes in inputs affect outputs?
- Search for an optimal system configuration?

Design of Simulation Experiments

- *Experimental design* traditionally refers to physical experiments
  - Origins in agriculture, laboratory experiments
- Can recycle most such traditional methods into simulation experiments
  - Will discuss some of this
- Also discuss different situation in simulation, both broader and more specific.

Design of Simulation Experiments

  - Overall purpose, what the outputs are, random-number use, effects of input changes on output, optimum-seeking
- Example questions in simulation experiment
  - What model configurations, versions to run?
- What are the input factors?
- How should they be varied?
- Use the same or different random numbers across configurations?
  - Run length?

Purpose of the Project?

- Maybe obvious, but be clear, specific about ultimate purpose of project
  - Answer can point different ways for design
  - Failure to ask/answer will leave you adrift – unlikely that you'll reach solid conclusions, recommendations
- Even if there's just one model in one configuration, or a very few fixed cases
  - Still questions on run length, number of runs, random-number allocation, output analysis

Purpose of the Project?

- But if there's more general interest in how changes in inputs affect outputs
  - Clearly, questions on which configurations to run
  - And all the single/few scenario questions above
  - Especially in optimum-seeking, need to take care in deciding which configurations to try, ignore
- Goals, strategies often become more ambitious (or less ...) during project
  - In designed experiments, can use results from early experiments to help choose later ones

Types of goals

Cycle	Goal
1 . Early	Validation
2 . Next	Screening
3 . Middle	Sensitivity Analysis, Understanding
4 . Middle	Predictive Models
5 . Later	Optimization, Robust Design

Output Performance Measures?

- Think ahead about what you want out of your simulations
- Most simulation software produces lots of default output
  - Time-based measures, counts
  - Economic-based measures (cost, value added)
  - can specify or create more
  - Often get averages, minima, maxima
- Easier to ignore things you have than to get
- things you don't have (to state the obvious …)
  - But extraneous output can significantly slow runs

Cont…d

- One fundamental question for output measures – time frame of simulation/system
  - *Terminating* (a.k.a. *transient*, *short-run*, *finitehorizon*)
- There’s a natural way to start and stop a run
- Start/stop rules set by system and model, not by you
- Need to get these right – part of building a valid model
  - *Steady-state* (a.k.a. *long-run*, *infinite-horizon*)
- Outputs defined as a limit as simulation run length $\rightarrow\infty$
- No natural way to start – system has already been running forever

Cont…d

- In stochastic simulation, outputs are observations from (unknown) probability distributions
  - Ideally, estimate the whole distribution – ambitious goal
- Usually get summary measures of output distributions
  - Means (maybe too much focus on these)
  - Extrema
  - Variance, standard deviation
  - Quantiles of output distribution
- Output desired can affect model, data structure

How to use Random number generation

- Most simulation models are *stochastic* – Random inputs from probability distributions
- Simulation software has ways to *generate* observations from input distributions
  - Rely on *random-number generator*
- Algorithm to produce a sequence of values that appear independent, uniformly distributed on [0, 1]
  - RNGs are actually fixed, recursive formulae generating the same sequence
  - Will eventually *cycle*, and repeat same sequence

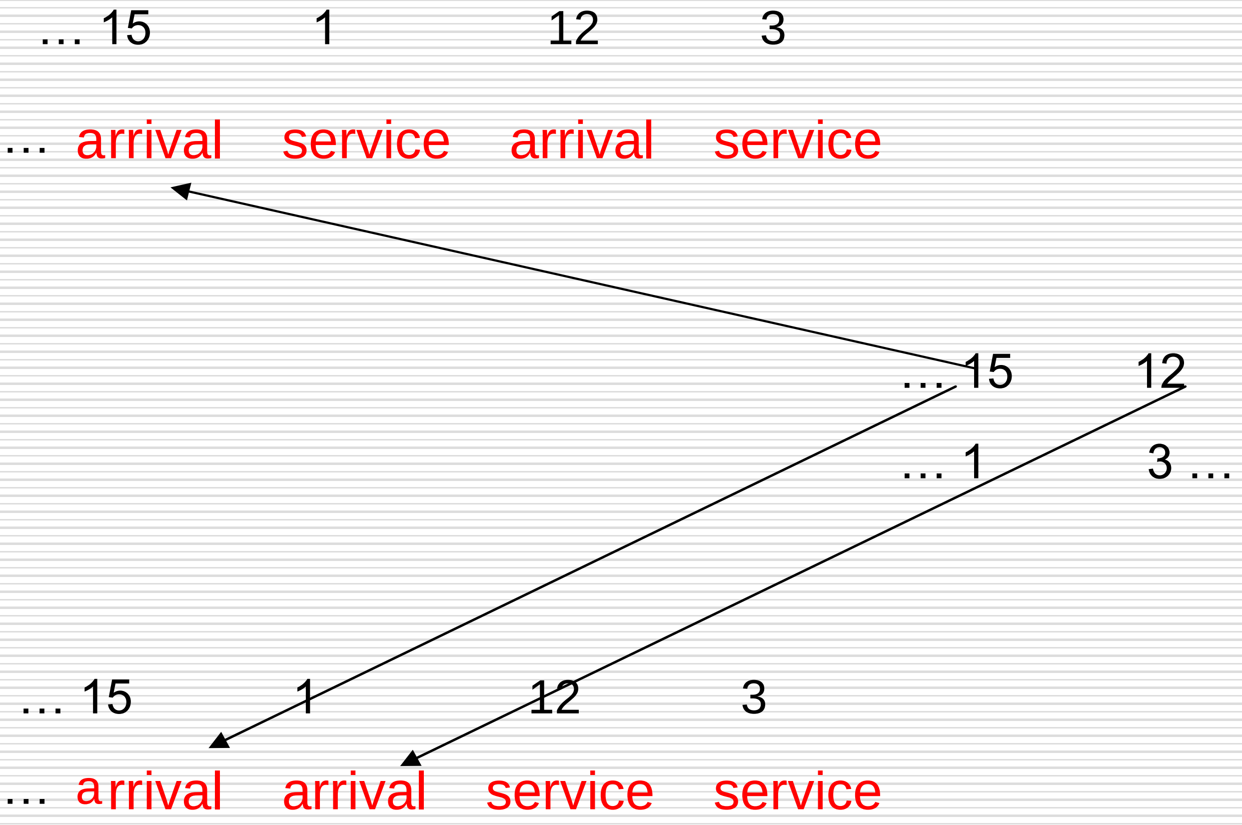
Random number generation..Cont

- RNG is controllable, so randomness in simulation experiment is controllable – useful?
  - Controlling carefully is one way to reduce variance of output, without simulating more
- Part of designing simulation experiments is to decide how to allocate random numbers – First thought – independent (no reuse) throughout
- Certainly valid and simple statistically
- But gives up variance-reduction possibility

Random number generation..Cont

- Usually takes active intervention in simulation software – New run always starts with same random numbers – override
- Better idea when comparing configurations
- Re-use random numbers across configurations – *common random numbers*
- Differences in output more likely due to differences in configurations, not because the random numbers bounced differently (they didn’t) – Probabilistic rationale:
- $\text{Var}(A - B) = \text{Var}(A) + \text{Var}(B) - 2\,\text{Cov}(A, B)$
- Hopefully, $\text{Cov}(A, B) > 0$ under CRN
- Usually true, though (pathological) exceptions exist
- Must ***synchronize*** RN use across configurations

Random number generation..Cont



Separate ‘arrival’ and ‘service’ streams

Sensitivity of Outputs to Inputs?

- Sensitivity of Outputs to Inputs?
  - Simulation models involve *input factors*
- Quantitative – arrival rate, number of servers, pass/fail probabilities, job-type percentages, ...
- Qualitative – queue discipline, topology of part flow, shape of process-time distribution, ...
- *Controllable* vs. *uncontrollable* input factors – In **real** system, usually have both

Number of servers, queue discipline – controllable

- Arrival rate, process-time-distribution – uncontrollable – In **simulation**, *everything* is controllable
- Facilitates easy "what-if" experimentation
- Advantage of simulation vs. real-world experimentation

Sensitivity of Outputs to Inputs?

- Input factors presumably have *some* effect on output – what kind of effect?
  – Sign, magnitude, significance, linearity, ...
- Mathematical model of a simulation model:

$$\text{Output}_1 = f_1(\text{Input}_1, \text{Input}_2, \ldots)$$
$$\text{Output}_2 = f_2(\text{Input}_1, \text{Input}_2, \ldots)$$
$$\vdots$$

f_1, f_2, ... represent simulation model itself

- Common goal – estimate change in an output given a change in an input
  – Partial derivative
  – But we don't know f_1, f_2, ... (why we're simulating)
  – Now discuss different estimation strategies

Classical Experimental Design

- Has been around for ~80 years
  - Roots in agricultural experiments
- Terminology
  - Inputs = Factors
  - Outputs = Responses
- Estimate how changes in factors affect responses
- Can be used in simulation as well as physical experiments
  - In simulation, have some extra opportunities
- Two-level factorial designs

Cont..d

- Each input factor has two levels ("–", "+" levels)
- No general prescription for setting numerical levels

- Should be "opposite" but not extreme or unrealistic
  - If there are k input factors, get 2^k different combinations of them … 2^k *factorial design*
  - Run simulation at each combination • Replicate it? Replicate whole design?
  - Get responses $R_1, R_2, \ldots, R_{2^k}$
  - Use to learn about effects of input factors

Classical Experimental Design Cont.

- Design matrix for k = 3 (with responses):

Run (i)	Factor 1	Factor 2	Factor 3	Response
1	–	–	–	R_1
2	+	–	–	R_2
3	–	+	–	R_3
4	+	+	–	R_4
5	–	–	+	R_5
6	+	–	+	R_6
7	–	+	+	R_7
8	+	+	+	R_8

- *Main effect* of a factor: average change in response when factor moves from "–" to "+"
  - Main effect of factor 2:

Classical Experimental Design Cont.

- $(-R_1 - R_2 + R_3 + R_4 - R_5 - R_6 + R_7 + R_8)/4$
- Two-way interaction: does the effect of one factor depend on the level of another?
  - "Multiply" sign columns of the two factors, apply to response column, add, divide by 2^{k-1} – Interaction between factors 1 and 3:
- $(+R_1 - R_2 + R_3 - R_4 - R_5 + R_6 - R_7 + R_8)/4$
  - If an interaction is present, cannot interpret main effects of involved factors in isolation

Classical Experimental Design Cont.

Example: car maintenance/repair shop

Kelton, Sadowski, Sturrock, *Simulation With Arena ,3rd ed.2004,*

Outputs:

Daily profit

Daily late wait jobs=cars/day that are late for customers waiting

Inputs:

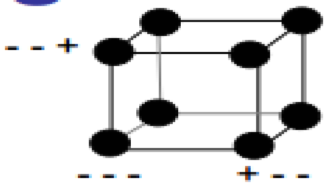
Max load=max hours/day that can be booked

Max wait=max number of customer-waiting cars/day that can be booked

Wait allowance=hours padded to predicted time in system for waiting customers.

Classical Experimental Design Cont.

- 2^3 factorial design
  - 100 replications per design point
  - Used Arena Process Analyzer to manage runs:



Scenario Properties				Controls			Responses	
S	Name	Program File	Reps	Max Load	Max Wait	Wait Allowance	Daily Profit	Daily Late Wait Jobs
	Base Case	1 : Model 06-04.p	100	24.0000	5.0000	1.0000	492.628	0.699
	- - -	1 : Model 06-04.p	100	20.0000	1.0000	0.5000	326.596	0.257
	+ - -	1 : Model 06-04.p	100	40.0000	1.0000	0.5000	489.135	0.259
	- + -	1 : Model 06-04.p	100	20.0000	7.0000	0.5000	328.400	0.892
	+ + -	1 : Model 06-04.p	100	40.0000	7.0000	0.5000	480.321	1.771
	- - +	1 : Model 06-04.p	100	20.0000	1.0000	2.0000	326.596	0.083
	+ - +	1 : Model 06-04.p	100	40.0000	1.0000	2.0000	489.135	0.076
	- + +	1 : Model 06-04.p	100	20.0000	7.0000	2.0000	328.400	0.284
	+ + +	1 : Model 06-04.p	100	40.0000	7.0000	2.0000	480.321	0.590

  - Main effects on Daily Profit: +157, –4, 0
    - Implication: should set Max Load to its "+" value
    - Other two factors don't matter
  - Interactions on Daily Profit: –5 (1x2), others 0

Link to spreadsheet

25

VRT

- Variance reduction techniques require additional computation in order to be implemented.
- **Not for sure** whether a variance reduction technique will effectively reduce the variance in comparison with straightforward simulation.
- Common practice is to carry out a plot experiment
- a) the antithetic variates technique and
- (b) the control variates technique.

The antithetic variates technique AVT

- is a very simple technique to use and it only requires a few additional instructions in order to be implemented
- No general guarantee of its effectiveness can be given
- Therefore, a small pilot study may be useful in order to decide whether or not to implement this technique.

Variance Reduction Techniques VRT

- The accuracy of an estimate is proportional to $1/\sqrt{n}$, where n is the sample size.
- One way to increase the accuracy of an estimate increase n
  - To halve the confidence interval 4n should be used
- BUT it requires long time and expensive (memory and cpu)
- An alternative way to increasing the estimate's accuracy is to reduce its variance.
  - If one can reduce the variance of an endogenously created random variable without disturbing its expected value, then the confidence interval width will be smaller, for the same amount of simulation
- Techniques aiming at reducing the variance of a random variable are known as ***Variance Reduction Techniques***

AVT

- Let X be an endogenously created random variable
  - $x_1^{(1)}, x_2^{(1)}, \ldots, x_n^{(1)}$ realizations of X obtained in a simulation run.
  - $x_1^{(2)}, x_2^{(2)}, \ldots, x_n^{(2)}$ observations of X obtained in a second simulation run.
- Now, let us define a new random variable

$$z_i = (x_i^{(1)} + x_i^{(2)})/2, \quad i = 1, 2, \ldots, n.$$

AVT cont.

- More specifically, let $Z=(X^{(1)}+X^{(2)})/2$, $X^{(i)}$, i=1,2, indicates the random variable X as observed in the ith simulation run. We have

$$E(Z)= E\left(\frac{X^{(1)}+X^{(2)}}{2}\right)$$

$$=\frac{1}{2}\,[E(X^{(1)})+E(X^{(2)})]$$

$$=E(X)$$

- Thus, the expected value of this new random variable Z is identical to that of X. Now, let us examine its variance. We have

AVT

$$\mathrm{Var}(Z) = \mathrm{Var}\left(\frac{X^{(1)}+X^{(2)}}{2}\right)$$

$$= \frac{1}{4}\ [\mathrm{Var}(X^{(1)}) + \mathrm{Var}(X^{(2)}) + 2\mathrm{Cov}(X^{(1)}, X^{(2)})].$$

Remembering that $\mathrm{Var}(X^{(1)})=\mathrm{Var}(X^{(2)})=\mathrm{Var}(X)$, we have that

$$\mathrm{Var}(Z) = \frac{1}{2}\,(\mathrm{Var}(X) + \mathrm{Cov}\,(X^{(1)}, X^{(2)}))\ .$$

Since $\mathrm{Cov}(X,Y)=\rho\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}$, we have that

$$\mathrm{Var}(Z) = \frac{1}{2}\ \mathrm{Var}(X)\ (1 + \rho),$$

where ρ is the correlation between X(1) and X(2)

AVT

- In order to construct an interval estimate of E(X), we use random variable Z.
- we can cause Var(Z) to become significantly less than Var(X).
- This is achieved by causing ρ to become negative.
- In the special case where the two sets of observations X1 and X2 are indipendent of each other we have that p=0 hence Var(Z)=Var(X)/2
- The antithetic variates technique attempts to introduce a negative correlation between the two sets of observations.

Example

- As an example, let us consider a simulation model of a single server queue, and let X and Y indicate the waiting time in the queue and the interarrival time respectively.
- If Y is very small, then customers arrive faster and, therefore, the queue size gets larger.
- The larger the queue size, the more a customer has to wait in the queue, i.e. X is larger.
- On the other hand, if Y is large, then customers arrive slower and, hence, the queue size gets smaller. Obviously, the smaller the queue size, the less a customer has to wait in the queue, i.e., X is small.
- Therefore, we see that X and Y can be negatively correlated.

Example Cont.

- This negative correlation between these two variables can be created in a systematic way as follows.
  - Let F(t) and G(S) be the cumulative distribution of the inter- arrival and service time respectively
  - Let ri and vi be pseudo-random numbers.
- Then, $t_i=F^{-1}(r_i)$ and $s_i=G^{-1}(v_i)$ are an interarrival and a service variate. These two variates can be associated with the ith simulated customer
- An indication of whether the queue is tending to increase or decrease can be obtained by considering the difference $d_i=t_i-s_i$.
- This difference may be positive or negative indicating that the queue is going through a busy or slack period respectively

Ex. Contrast

- Now, let us consider that in the second run, we associate pseudo-random number r'i and v'i with the ith simulated customer, so that below equation has the opposite sign of di.

$$d_i' = t_i' - s_i' \text{ (where } t_i' = F^{-1}(r_i') \text{ and } s_i' = G^{-1}(v_i'))$$

- That is, if the queue was going through a slack (busy) period in the first run at the time of the ith simulated customer, now it goes through a busy (slack) period
- It can be shown that this can be achieved by simply setting r'i = 1-ri and v'i =1-vi.

Cont..d

- We make use of two controllable variables, Y1 and Y2, indicating the interarrival time and the service time respectively.
- These two random variables are strongly correlated with X, the waiting time in the queue.
- Yj(1) and Yj(2), j=1,2 can be negatively correlated by simply using the compliment of the pseudo-random numbers used in the first run.

Implementing technique

- Simulate the single server queue, and let , $x_1^{(1)}, x_2^{(1)}, \ldots, x_n^{(1)}$ be *n* i.i.d observations of X. (Independent and identically distributed
- Re-run the simulation, thus replicating the results, using pseudo-random numbers (ri,vi)=(1-r,1-vi). Let $x_1^{(2)}, x_2^{(2)}, \ldots, x_n^{(2)}$ be realization of X. Construct the interval estimate of E(X) using random variable z as described above.
- Obviously, the correlation between the two samples of observations is as good as the correlation between *Y1j* and *Yj*2, j=1,2.

- implemented in simulation of an M/M/1 queue
- The random variable X is the time a customer spends in the system
- The i.i.d. observations of X were obtained by sampling every 10th customer.

Sample size n	Confidence interval
600	13.86 ± 3.46
900	13.03 ± 2.70
1200	13.11 ± 2.30
1500	12.82 ± 1.99
1800	12.86 ± 1.84

Table 7.1: Straight simulation of an M/M/1 queue.

- Using the antithetic variates technique, we obtained a confidence interval of 13.52±1.76
- antithetic variates techniques were employed using two sets of observations each of size equal to 300, i.e., a total of 600 observations.

Implementing technique cont..d

- Simulate the single server queue, and let , $x_1^{(1)}$, $x_2^{(1)}$, ..., $x_n^{(1)}$ be n i.i.d observations of X. (
- Re-run the simulation, thus replicating the results, using pseudo-random numbers (ri,vi)=(1-r,1-vi). Let $x_1^{(2)}$, $x_2^{(2)}$, ..., $x_n^{(2)}$ be realization of X. Construct the interval estimate of E(X) using random variable z as described above.
- Obviously, the correlation between the two samples of observations is as good as the correlation between *Y*1*j* and *Yj*2, j=1,2.

Implementing technique cont..d

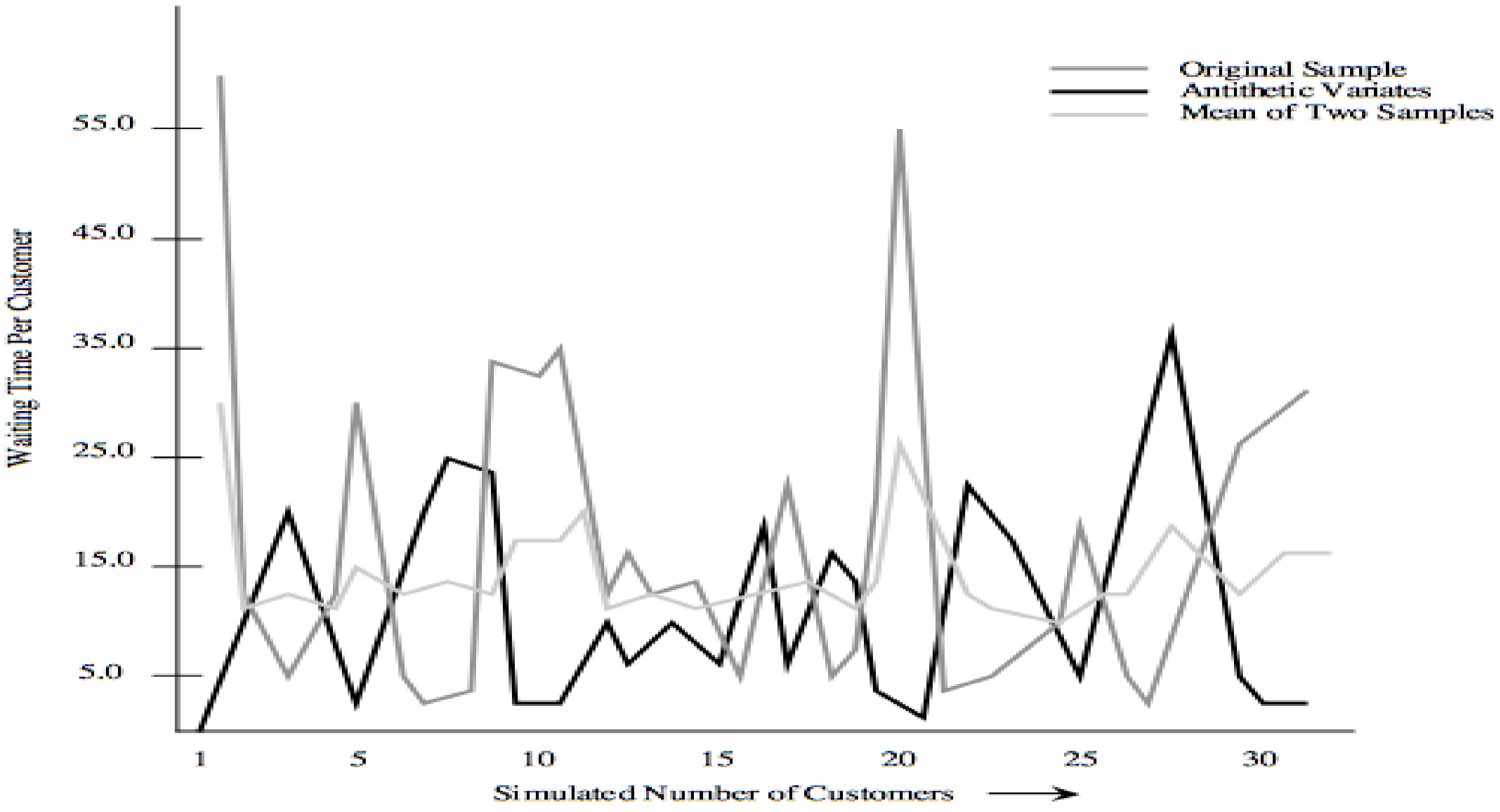
- 

Figure 7.1: Antithetic variates technique applied to an M/M/1 queue.

Implementing technique cont..d

- In the above example, the antithetic variates technique worked quite well. However, this should not be construed that this method always works well.
- In particular, in the following example, an M/M/2 queuing system was simulated

Type of Simulation	Sample Size	Mean	Standard Dev.	Conf. Interval	Standard Error	Standard Error as a % of Mean	Cost
Straight	400	18.85	9.70	+/-3.31	1.69	8.96	$.68
Straight	800	16.00	8.60	2.07	1.06	6.61	.72
Standard Antithetic	800	17.23	6.73	2.30	1.17	6.79	.97
Straight	1600	16.04	9.67	1.64	0.84	5.22	.79
Standard Antithetic	1600	15.69	5.98	1.44	0.74	4.69	1.03
Straight	2400	16.48	9.82	1.36	0.69	4.21	.88
Standard Antithetic	2400	15.84	6.36	1.25	0.64	4.01	1.09
Straight	3000	15.92	9.59	1.19	0.61	3.81	.93
Standard Antithetic	3000	16.11	6.96	1.17	0.60	3.72	1.12

Table 7.3: Straight simulation and antithetic variates techniques for an M/M/2 queue.

Implementing technique cont..d

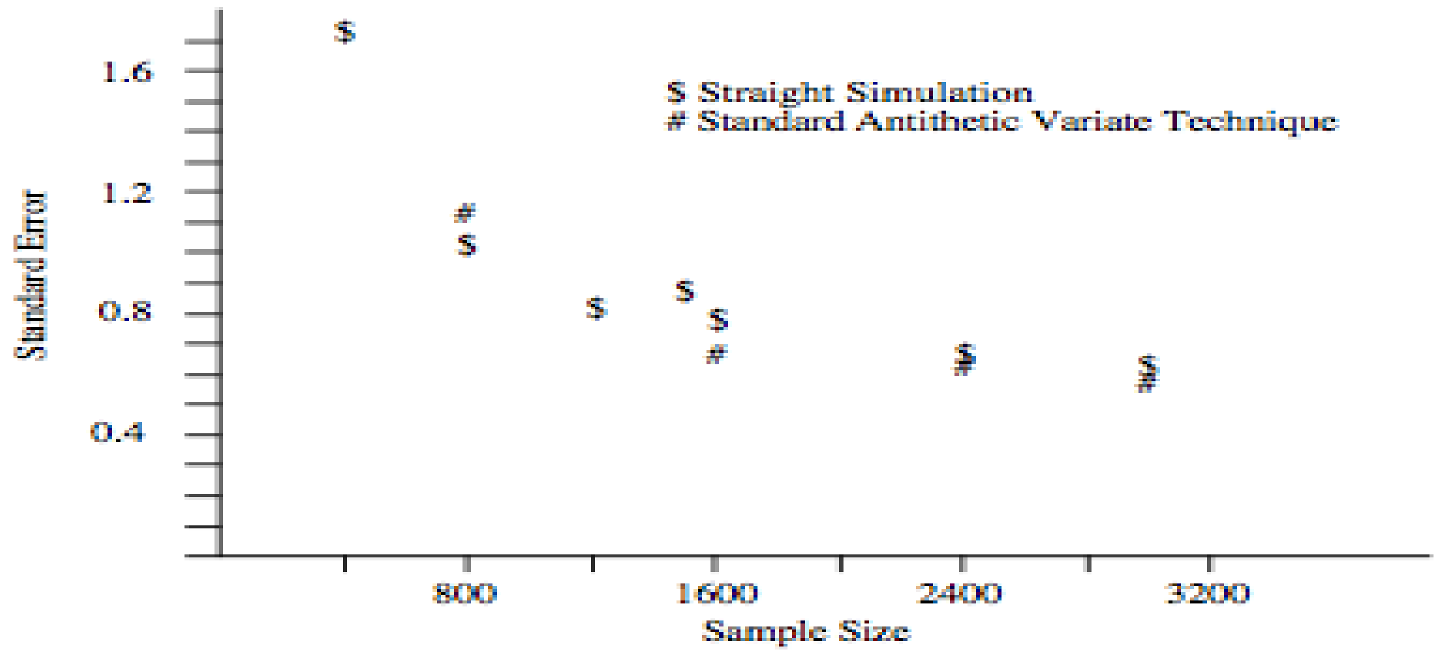
- 

Figure 7.2: Standard error plotted against sample size for straight simulation and antithetic variates techniques for an M/M/2 queue.

The control variates technique

- This method is otherwise known as the method of *Concomitant Information*.
- Let X be an endogenously created random variable whose mean we wish to estimate.
- Let Y be another endogenously created random variable whose mean is known in advance known as the *control* variable.
- Random variable Y is strongly correlated with X.

The control variates technique cont..d

a. X and Y are negatively correlated:

Define a new random variable Z so that $Z=X+Y-E(Y)$. We have

$$E(Z)=E(X+Y-E(Y))=E(X),$$

$$Var(Z) = Var(X) + Var(Y) + 2\,Cov(X, Y).$$

Since X and Y are negatively correlated, we have that $Cov(X,Y)<0$. Therefore, if $Var(Y) - 2\,|Cov(X,Y)| < 0$ then, a reduction in the variance of Z has been achieved.

Do more…