*Proceedings of the 2024 Winter Simulation Conference*
*H. Lam, E. Azar, D. Batur, S. Gao, W. Xie, S. R. Hunter, and M. D. Rossetti, eds.*

# TWENTY-THREE CRITICAL PITFALLS IN SIMULATION MODELING AND HOW TO AVOID THEM

Averill M. Law

Averill M. Law & Associates, Inc., 4955 East Calle Guebabi, Tucson, AZ, USA

## ABSTRACT

Many simulation projects are less than successful because "analysts" view simulation modeling as a complicated exercise in computer programming. This is probably caused by their education being limited to vendor training or an undergraduate simulation course that focuses on how to use a particular simulation-software package. Unsuccessful projects also result from lack of real-world experience in performing simulation studies. In this tutorial we discuss 23 critical pitfalls that can cause a simulation project to result in failure. These pitfalls fall into four categories: (1) modeling and validation, (2) simulation software, (3) modeling system randomness, and (4) design and analysis of simulation experiments.

## 1 INTRODUCTION

In many "simulation studies" a great amount of time and money is spent on model development and programming but little effort is made to address fundamental methodological issues such as model validation, selecting input probability distributions, and using correct statistical techniques to design and analyze simulation experiments. (See section 1.6 in Law 2024 for the steps in a sound simulation study.) These omissions are probably due to an analyst's education being limited to vendor training or an undergraduate course on simulation modeling that focuses on how to use a particular simulation-software package. Even if the analyst has taken a comprehensive graduate-level simulation course, there are still many opportunities for failure due to lack of real-world experience in performing simulation studies. Critical project-management ideas like the importance of a definitive problem formulation and regular interaction with management are typically not taught in university courses, and have to be "learned on the job."

In this tutorial we discuss 23 pitfalls to project success that were obtained from the following sources:

- Approximately 50 real-world simulation studies that we performed over many years
- The successes and failures of the 7000-8000 students who have attended my public and onsite simulation courses
- Reading hundreds of technical papers and books on simulation modeling

These pitfalls fall into four categories: (1) modeling and validation, (2) simulation software, (3) modeling system randomness, and (4) design and analysis of simulation experiments. For each potential pitfall, we give a brief discussion of how it can be avoided in practice and also a reference to more extensive information. Real-world examples of many of the important pitfalls are also presented.

## 2 MODELING AND VALIDATION

***Pitfall Number 1:*** Failure to have a well-defined set of objectives at the beginning of the study

There needs to be an initial kick-off meeting to determine the overall objectives of the study, *specific questions to be addressed*, performance measures of interest, scope of the project, etc. The meeting needs to be attended by all important stakeholders, including the project manager, subject-matter experts (SMEs), and simulation analysts. It is not sufficient to know just the overall objectives of the study, but rather one needs to have a list of the specific questions that the study needs to answer as well as the performance measures that the model needs to produce. Otherwise, it will be impossible to decide on an appropriate level of model detail or fidelity.

> **Example 1.** At a kick-off meeting for a study concerning the loading and transportation of crude oil in Alaska, the only people present were the "corporate champion" and two of us analysts. Other important stakeholders were not present due to the large geographical distances involved. As a result the "assumptions document" (see Pitfall Number 8 and section 5.4.3 in Law 2024) initially had many missing and incorrect assumptions. This problem was addressed by a structured walk-though of the assumptions document as discussed in Example 7.

***Pitfall Number 2:*** Misunderstanding of simulation by management

This can result in confusion with regard to the types of problems that are addressable, the time required for a sound study, and the expertise needed. A one-hour seminar presented by an analyst might help.

> **Example 2.** We did a simulation study for a major manufacturer of candy pieces to evaluate the efficacy of the packaging area. After the study was successfully completed, the plant manager insisted that we write a proposal to determine why the candy pieces were cracking during the production process. This was not a legitimate use of discrete-event simulation, caused by a lack of understanding on the part of management.

***Pitfall Number 3:*** Failure to communicate with management on a regular basis

This helps ensure the correct problem is being solved, the manager's interest is maintained, and, importantly, that the manager understands and agrees with the key model assumptions. The greatest model for the wrong problem will be of little value. If the manager does not *understand and agree with the key assumptions*, "The model is likely to be put on a shelf somewhere, only to collect dust."

> **Example 3.** A military analyst worked on a project for months without interacting with the requesting general. At the final briefing in the U.S. Pentagon, the general walked out after five minutes stating, "That's not the problem I'm interested in."

***Pitfall Number 4:*** Failure to collect good system data

Data are needed for specifying input probability distributions (see Pitfall Number 17 and chapter 6 in Law 2024) and for model validation (see Pitfall Number 10 and section 5.4 in Law 2024). Note that validation is the *process* of determining whether a simulation model is an accurate representation of the system, for the *particular objectives of the study*.

***Pitfall Number 5:*** Inappropriate level of model detail

There should *not* be a one-to-one correspondence between each model element and each system element. Start with a "simple" model and embellish it as needed. The adequacy of a particular version of the model is assessed by having the model reviewed by managers and SMEs. A common pitfall of beginning modelers is to put every detail of the system into the model to "guarantee" its validity. In general, a model should have just enough detail to answer the questions of interest.

> **Example 4.** A system designed to produce pet food consisted of a meat plant and a cannery. In the meat plant, meat was either ground fine or into chunks and then placed into buckets and transported to the cannery by an overhead conveyor system. In the cannery, buckets were dumped into

mixers that process the meat and then dispense it to the filler/seamers for canning. The empty buckets are conveyed back to the meat plant for refilling.

Originally, it was decided that the system producing the chunky product was relatively unimportant and, thus, it was modeled in a simple manner. However, at the structured walk-through of the model (see Pitfall Number 9 and section 5.4.3 in Law 2024), machine operators stated that this subsystem was actually much more complex. To increase model validity and gain credibility with these members of the project team, we had to include machine breakdowns and contention for resources. Furthermore, after the initial runs of the model were made, it was necessary to make additional changes to the model suggested by a mixer operator.

***Pitfall Number 6***: Treating a simulation study as if it were a complicated exercise in computer *programming*

In some studies, the first thing that the organization considers is what simulation software to purchase or use. However, this decision should only be made after the problem of interest has been precisely stated, in order to ensure the software's appropriateness for the problem at hand. Moreover, in general, "programming" is only 25 to 50 percent of a sound study.

***Pitfall Number 7:*** Lack of knowledge of simulation methodology and probability/statistics

This is caused by a modeler's education being limited to vendor training or university courses that focus on the use of a particular software product. A knowledge of model validation, selecting input probability distributions (see Section 4), basics of random-number generators, design and analysis of simulation experiments (see Section 5), and project management is critical to a project's success.

**Example 5.** I presented an onsite seminar on simulation for a defense contractor. At the end of the first day, the manager, who also attended the seminar, said that the attendees wanted to know why they needed to understand probability and statistics to do simulation modeling. In amazement, I asked the manager what background they required of a job candidate to be hired as a simulation modeler. He said, "They must have taken a university course on their preferred simulation product."

***Pitfall Number 8:*** Failure to thoroughly document the simulation model

The analyst needs to create a detailed assumptions document (see section 5.4.3 in Law 2024) with the following elements:

- Overall objectives, specific questions that need to be answered, model inputs, and performance measures for the simulation project
- Process-flow/system-layout diagram
- Detailed descriptions of each subsystem and their interactions
- Simplifying assumptions that were made and their justification
- Summaries of data
- Limitations of the model
- Sources of important or controversial information

The assumptions document should be written in bullet format for easy review at meetings. Also, the assumptions document should be a "blueprint" for creating the simulation program. It is more detailed than what some people call a conceptual model and it is also different from a requirements document.

**Example 6.** We were asked to help validate a simulation model for a major military weapons system, which would cost billions of dollars. The model had been under development for ten years and consisted of more than 10,000 lines of code in an outdated simulation-software package. Moreover, the documentation for the model primarily consisted of 35 PowerPoint slides, making it extremely difficult to know what assumptions had been made. In conversations with the model developer, it was found that some model assumptions were, in fact, wrong and that the methodology used to "validate" the model was suspect. Despite these shortcomings the model was officially "accredited" for use.

***Pitfall Number 9:*** Failure to perform a structured walk-through of the assumptions document

The assumptions document should be formally reviewed before an audience of all stakeholders, by going from one bullet to the next. A key benefit is the resulting interactions among the subject-matter experts who are present. Model errors and omissions will be found! They should be corrected before proceeding with programming. In advance of the formal structured walk-through, a copy of the assumptions document should be sent to all of the project stakeholders and their feedback requested. Use any worthwhile information obtained to update the documentation. However, these people will be busy doing their own daily work and may not have the time or inclination to take this seriously. Therefore, this informal review is probably not sufficient (see section 5.4.3 in Law 2024).

**Example 7.** *(continuation of Example 1)* This simulation study was precipitated by the *Exxon Valdez* oil spill of 1989 that occurred in Alaska, where an oil tanker of the same name ran aground on a reef resulting in millions of gallons of crude oil being spilled and considerable harm to wildlife. As a result, the U.S. Congress passed the Oil Protection Act of 1990 that mandated that all existing tankers had to be replaced by a double-hull design by no later than 2015.

<u>System of interest:</u>

Crude oil was loaded onto tankers at a port in Alaska and transported down the west coast of the U.S. to refineries in the states of Washington and California. The operation of the port is greatly affected by severe weather.

<u>Simulation study:</u>

We were contacted by a major oil company in November 1997 to build a simulation model for determining the number of double-hulled tankers required for the system described previously. An initial estimate of the number of required tankers, which cost 200 million dollars each, was obtained by a spreadsheet analysis and a simulation model was desired for confirmation.

The assumptions document had sections corresponding to the following subsystems:

- The extraction of crude oil from the Alaskan North Slope (simplified model)
- The loading of oil onto tankers at the port
- The unloading and storage of oil at the refineries
- Nature and effect of weather
- Maintenance of tankers in Asia

Much of the information for the assumptions document was obtained from a two-day kickoff meeting in Alaska during November 1997, which involved two of us analysts and a *single* "corporate champion" from the oil company. Unfortunately, crucial SMEs were not present due to the large geographical distances involved!

<u>Structured walk-through:</u>

The structured walk-through took place in California during January of 1998. *Many* of the important model assumptions were found to be incorrect or missing, due to the lack of critical SMEs at the kickoff meeting. For example, it was discovered that the effect of severe weather on the port operations was much more complicated than we were previously told. Also, we found out the real disposition of excess oil when a tanker arrived at a refinery with inadequate remaining storage. Namely, the excess oil was sold off to a competing company.

As a result of the disappointing structured walk-through, various SMEs at the meeting were given the responsibility of gathering information on different parts of the system, and they provided the required information to us within two weeks. The assumptions document was then updated, the simulation analysis was performed, and a second walk-through was successfully performed at the final presentation for the study. The simulation study was purported to have saved the oil company 52 million dollars.

***Pitfall Number 10:*** Failure to validate the simulation model

The most definitive validation technique is to first simulate an existing system configuration (if any) that is similar to the system of interest, and then compare this model's output data to those from the existing system itself. If the two sets of data compare "closely," then the model of the *existing system configuration* is considered "valid." The model is then modified so that it represents the proposed system configuration. The greater the commonality between the existing and proposed systems, the greater is our confidence in the validity of the model of the *proposed system*.

**Example 8.** A system for producing paper pulp had two machines, but local management wanted to buy a third machine. A model was first built for the existing system, and model and system throughputs for the two machines differed by only *0.4 and 1.1 percent*, respectfully. A third identical machine was then added to the model but was found *not* to be necessary to meet system performance requirements. However, the key to the *actual use* of the model's results was the "buy-in" by the corporate vice president of manufacturing, resulting from the success of the validation exercise. See section 5.45 in Law 2024 for additional examples of this technique.

## 3 SIMULATION SOFTWARE

***Pitfall Number 11:*** Inappropriate simulation software

If significant programming is required to build a simulation model, then simulation may never actually be used and its potential benefits lost. Many, if not most, people who perform simulation studies have backgrounds in operations research and may not have proficiency in a general-purpose programming language such as Python or Java. However, most models will, in fact, require programming in some form despite the claims of some simulation-software vendors. On the other hand, if the software has inadequate modeling flexibility, then it may not be viable for more-complicated future projects.

***Pitfall Number 12:*** Belief that so-called "easy-to-use software" requires a lower level of technical competence

Simulation-software vendors often advertise how easy their software is to use. However, such software does not alleviate the need-to-know stochastic modeling (e.g., queueing theory), simulation methodology (model validation, selection of input probability distributions, design and analysis of simulation experiments, etc.), and probability/statistics.

***Pitfall Number 13:*** Lack of quality software documentation

There needs to be a university-quality textbook, specification of the logic for all modeling constructs, and *numerous detailed examples* with precise problem statements. Small models that illustrate the use of particular software features are also valuable.

***Pitfall Number 14:*** Misuse of animation

Animation is particularly important for model communication but is not a substitute for properly designed and analyzed statistical experiments. Three-dimensional animation is preferable because of the added realism that it provides. Animation is also useful for model debugging and for understanding the dynamic behavior of a system.

**Example 9.** A simulation model was developed for a candy manufacturing system. A newly promoted operations manager, who had no familiarity with the simulation model, declared, “That’s my system!” upon seeing an animation of his system for the first time – the model gained instant credibility.

***Pitfall Number 15:*** Assuming the statistical capabilities in simulation software are always correct

Some simulation software use methods for analyzing simulation output data that are not statistically valid (see Pitfall Number 19 and example 4.28 in Law 2024). Also, a number of simulation products use outdated or limited random-number generators. For example, a random-number generator (chapter 7 in Law 2024) should have a large number of distinct random-number streams, which is critical when comparing alternative system configurations (see chapter 10 and section 11.2 in Law 2024).

## 4 MODELING SYSTEM RANDOMNESS

***Pitfall Number 16:*** Replacing a probability distribution by its mean

This might be done because there are little or no data to determine an appropriate probability distribution or because the so-called analyst does not realize that this is a significant issue.

**Example 10.** Consider a single-server queueing system with interarrival times having an exponential distribution with mean 1 minute and service times having an exponential distribution with mean 0.99 minute. Then we have what is called the $M/M/1$ queue in the queueing-theory literature (see Ross 2019). Furthermore, it can be shown for this system that the long-run (steady-state) average number of customers in the queue is approximately 98. Suppose because of lack of knowledge, the analyst makes every interarrival time equal to 1 minute (the mean) and every service time equal to 0.99 minute (the mean). Then it should be clear that there are never any customers in the queue because each customer arrives to find that the previous one left the system 0.01 minute ago. Thus, one must also capture the variability of the input processes rather than just their means.

***Pitfall Number 17:*** Incorrect choice of input probability distributions

Simulation modelers often choose their input probability distributions to be normal or uniform because of familiarity (the normal distribution is ubiquitous in statistics books) or simplicity. However, these distributions are rarely appropriate for discrete-event simulation models. In Figure 1 we present a histogram of 910 processing times on a machine for an automotive manufacturer. Note that the histogram is skewed to the right (i.e., it has a longer right tail than left) and is shifted to the right of the origin by about 13 time units. In Figure 2 we give a histogram of 219 interarrival times of cars. We see that the histogram is once again skewed to the right and has an exponential-like shape.

We now show that the distribution used to model a source of randomness can, in fact, greatly affect the simulation results.

**Example 11.** A single-server queueing system (e.g., a single machine in a factory) has exponential interarrival times with a mean of 1 minute. Suppose that 200 service times are available from the system, but their underlying probability distribution is unknown. Using an approach that is discussed

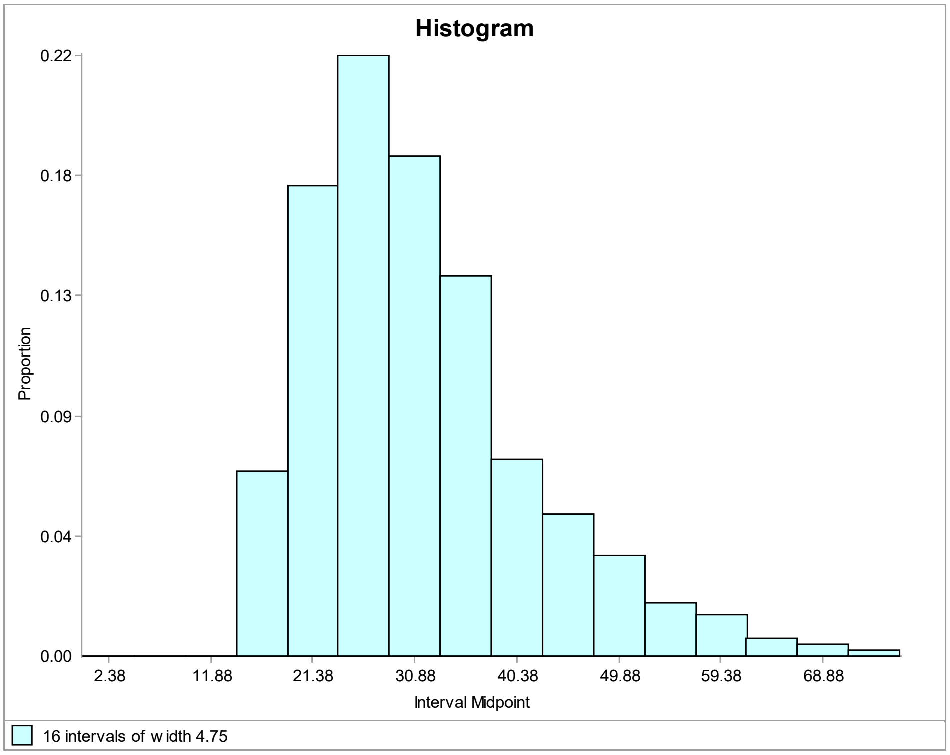


Figure 1: Histogram of 910 processing times for an automotive manufacturer.

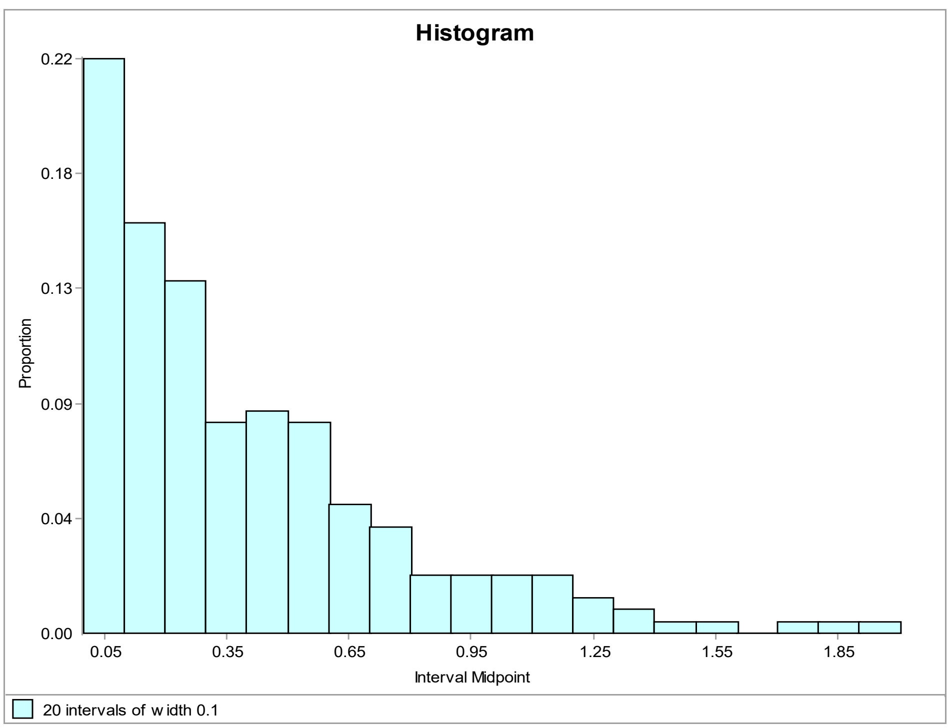


Figure 2: Histogram of 219 interarrival times to a drive-up bank.

in section 6.5 of Law (2024), we "fit" the best exponential, gamma, Weibull, lognormal, and normal distributions to the observed service-time data. We then made 100 independent simulation runs (i.e., different random numbers were used for each run, as discussed in section 7.2 of Law 2024) of the queueing system, using *each* of the five fitted distributions. (For the normal distribution, if a service time was negative, it was generated again.) Each of the 500 simulation runs was continued until 1000 delays in queue were collected. A summary of the results from these simulation runs is given in Table 1. We performed a thorough statistical analysis of the data and found that the Weibull distribution actually provides the best model for the service-time data. Thus, the average delay for the real system should be close to 4.36 minutes. On the other hand, the widely-used normal distribution gives an average delay of 6.04 minutes, corresponding to a shocking model output error of 39 percent!

***Pitfall Number 18:*** Cavalier use of the triangular distribution to represent a task time when data could be collected

The triangular distribution is often used to represent a task time when there are no corresponding system data available or there is not time to collect the required amount of data. However, if there is time to collect data, then this should definitely be done, because, as we will see, the triangular distribution can perform quite poorly in practice. To define a triangular distribution for a task time $X$, subject-matter experts are asked for their subjective estimates of the following parameters:

$a$ = minimum task time

$b$ = maximum task time

$m$ = most-likely task time (the mode)

Then a triangular density function $f(x)$ with mode $m$ is placed on the interval $[a,b]$, as shown in Figure 3. One problem with using a triangular distribution is choosing the value for $b$. For example, is $b$ the maximum over the next three months or the maximum over a lifetime? Another problem is that the triangular distribution will probably not work well if the true underlying distribution has a long and thick right tail, which is illustrated in Example 12.

**Example 12.** Consider a single-server queueing system with exponential interarrival times with mean 1 and lognormal service times with mean 0.9, variance 1.39, and mode $m = 0.2$, as shown in Figure 4. However, the service-time distribution is actually *unknown* to the analyst, so he tries to approximate this distribution by a triangular distribution with $a = 0$, $m = 0.2$, and $b = 1.97$ (90th percentile for the lognormal), which is also shown in Figure 4 (in green). (The values for $a$ and $m$ are correct.) This triangular distribution is a reasonable approximation near the mode $m$, but it does quite poorly in the right tail. Alternatively, suppose the analyst tries to approximate the unknown lognormal distribution by a triangular distribution with $a = 0$, $m = 0.2$, and a mean of 0.9 (correct). Since the mean of any triangular distribution is $(a + b + m)/3$, this implies that $b = 2.5$. This triangular distribution is also shown in Figure 4 (in yellow). It does not do as well as the first

Table 1: Simulation results for the five service-time distributions (in minutes).

| Service-time distribution | Average delay in queue |
|---|---|
| Exponential | 6.71 |
| Gamma | 4.54 |
| Weibull | 4.36 |
| Lognormal | 7.19 |
| Normal | 6.04 |

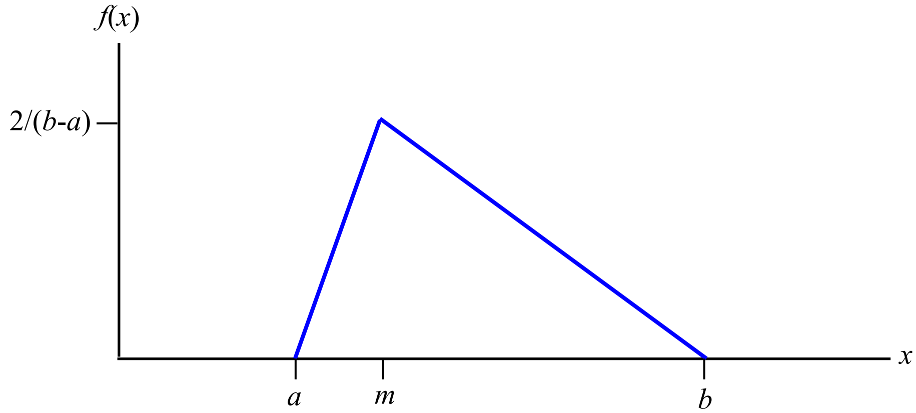


Figure 3: Triangular density function $f(x)$ on the interval $[a, b]$ with mode $m$.

near the mode but appears to do better in the right tail. For a more quantitative discussion of just how bad these triangular distributions are as approximations for the service-time distribution, see example 6.23 in Law 2024.

In summary, we recommend collecting data on all possible sources of system randomness, if at all possible.

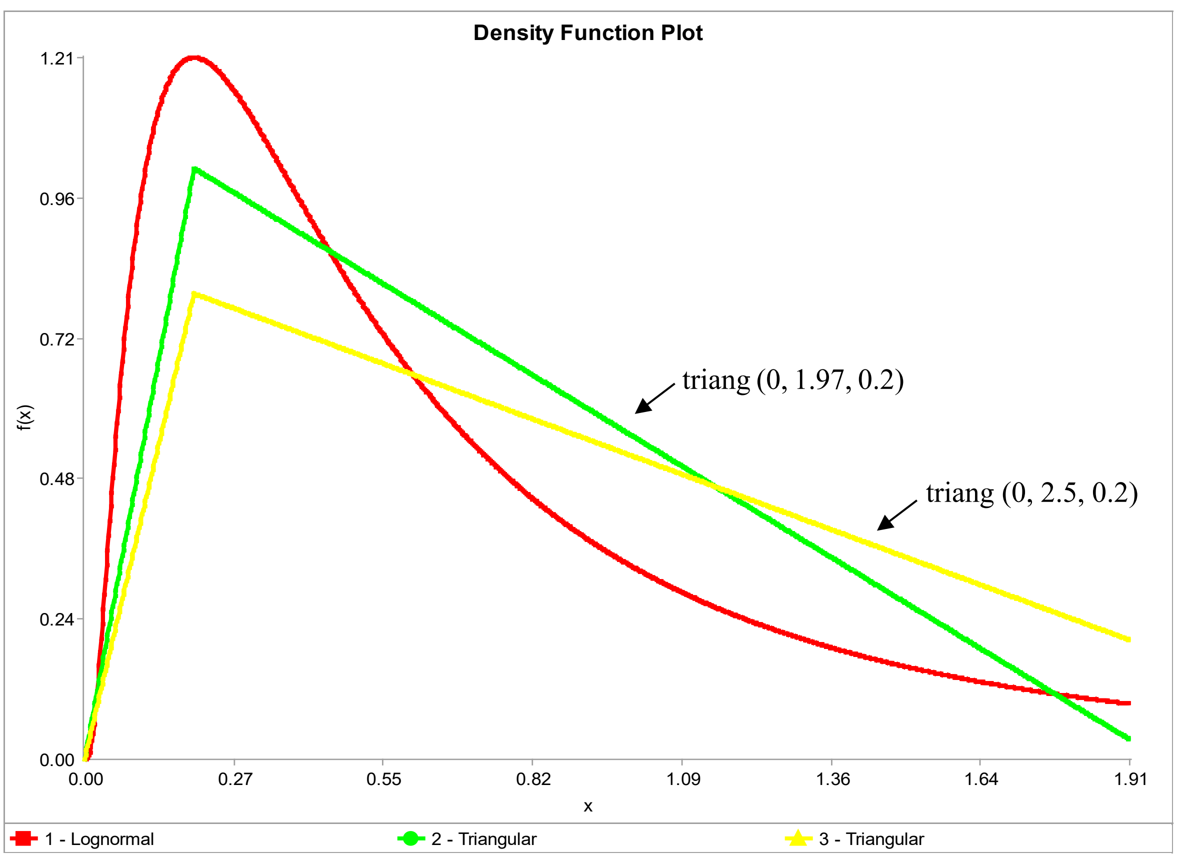


Figure 4: Lognormal and approximating triangular distributions.

## 5 DESIGN AND ANALYSIS OF SIMULATION EXPERIMENTS

***Pitfall Number 19:*** Using incorrect statistical formulas to analyze the output data from one simulation run

One should not analyze simulation output data from one run using equations and expressions from classical statistics that assume independent, identically distributed data, which will virtually never be the case in practice. Use of these equations will result in variance estimates that are biased low and a confidence interval whose actual coverage probability is drastically smaller than the nominal value (see example 4.28 in Law 2024). Unfortunately, one or more leading simulation packages apply these equations directly to simulation output data obtained from one run. See chapter 9 in Law 2024 for a comprehensive discussion of simulation output-data analysis.

***Pitfall Number 20:*** Making a single run of a particular system design and treating the output statistics as the "true answers"

The following example shows how dramatically simulation results can vary from one run to another because each run uses different U(0,1) random numbers.

> **Example 13.** Consider a department of motor vehicles (DMV) with five clerks that are being fed by a single queue. The DMV is open from 9 A.M. to 5 P.M. but will serve all customers present at closing. Assume that the interarrival times are exponentially distributed with mean 1 minute and service times are exponentially distributed with mean 4.5 minutes. We made 10 independent runs of the simulation with the results shown in Table 2. Note on run1 that 494 customers were served and their average delay in queue was 10.37 minutes. On the other hand, for run 5 there were 436 customers served and they had an average delay in queue of 1.83 minutes, which is approximately 18 percent of what it was on run 1.

***Pitfall Number 21:*** Failure to have a warmup period when doing a steady-state analysis

> **Example 14.** Consider a manufacturing system for which we want to estimate the steady-state (long-run) mean hourly throughput. The hourly throughputs will tend to be small at the "beginning" of a simulation run because it is typically assumed that there are no jobs present initially. The *warmup period* is the amount of simulation output data to delete until the mean hourly throughputs are representative of the long-run behavior of the system. In Figure 5 we show a plot produced by Welch's method (see pp. 407-412 in Law 2024), where the warmup period was chosen to be 120 hours since the average hourly throughput has approximately leveled off from hour 121 on. The warmup period should be chosen conservatively to make sure that all biased data have been deleted

Table 2: Results from 10 replications of the DMV.

| Replication | Number of customers served | Average delay in queue |
|---|---|---|
| 1 | 494 | 10.37 |
| 2 | 464 | 2.17 |
| 3 | 464 | 8.03 |
| 4 | 491 | 7.93 |
| 5 | 436 | 1.83 |
| 6 | 488 | 4.83 |
| 7 | 487 | 4.85 |
| 8 | 492 | 6.86 |
| 9 | 506 | 6.46 |
| 10 | 462 | 3.04 |

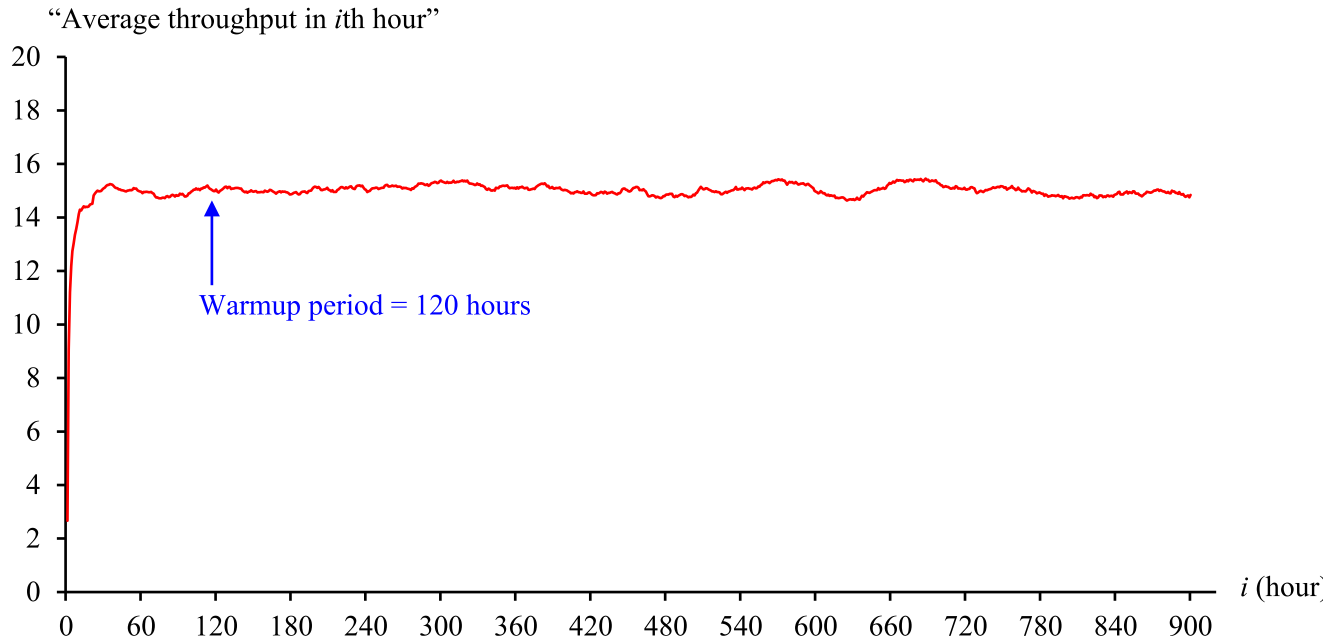


Figure 5: Welch plot for hourly throughputs.

from the beginning of the simulation run.

***Pitfall Number 22:*** Comparing two alternative system designs without a sound statistical approach often using only one run of each

See sections 10.2.1 and 11.2 in Law 2024 for a discussion of how a confidence interval can be used to compare the means of two competing system designs.

***Pitfall Number 23:*** Using the wrong performance measure to evaluate system performance

**Example 15.** Consider the two configurations of a DMV with three clerks shown in Figure 6. The DMV on the left has three clerks each with its own queue. However, a customer will jockey from one queue to another if she thinks that it is to her advantage. Namely, she will jockey if there are two more customers in her column (in service plus in queue) than the number of customers in another clerk's column. It takes no time to jockey. The DMV on the right has three clerks with one queue feeding them all.

For which configuration will the average delay in queue of a customer be the smallest? Perhaps surprisingly, it turns out that the average delay is the same for both configurations. (The average

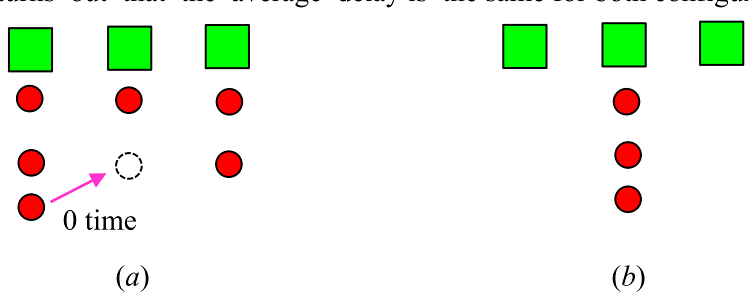


Figure 6: Two DMVs with three clerks: (*a*) three queues and jockeying, (*b*) one queue.

number in queue is also the same.) Thus, if one evaluates the efficacy of the two configurations based solely on the average delay, the conclusion is that there is no difference. On the other hand, the variability of the delay in queue is smaller for the one-queue configuration. This fact along with the greater equity of the one-queue policy, has probably led many organizations, e.g., DMVs and airlines, to adopt this policy. See example 9.20 in Law 2024 for further discussion.

## REFERENCES

Law, A. M. 2024. *Simulation Modeling and Analysis*, 6th ed., McGraw-Hill, New York.

Ross, S. M. 2019. *Introduction to Probability Models,* 12th ed., Academic Press, San Diego.

## AUTHOR BIOGRAPHY

**AVERILL M. LAW** is President of Averill M. Law & Associates, Inc., a company specializing in simulation seminars, simulation consulting, and software. He has presented more than 550 simulation and statistics short courses in 20 countries, including onsite seminars for AT&T, Australian Department of Defence, Boeing, Caterpillar, Coca-Cola, GE, GM, IBM (U.S., Belgium), Intel, Lockheed Martin, Los Alamos National Lab, NASA, NATO (Netherlands), NSA, Raytheon, Sandia, Sasol Technology (South Africa), 3M, UPS, U.S. Air Force, U.S. Army, U.S. Navy, and Verizon. He has written or coauthored numerous papers and books on simulation, operations research, statistics, manufacturing, and communications networks, including the book *Simulation Modeling and Analysis* that has been cited more than 24,800 times and is widely considered to be the “bible” of simulation. He developed the ExpertFit® distribution-fitting software and also several videotapes on simulation modeling. He was awarded the INFORMS Simulation Society Lifetime Professional Achievement Award in 2009. Dr. Law wrote a regular column on simulation for *Industrial Engineering* magazine. He has been a tenured faculty member at the University of Wisconsin-Madison and the University of Arizona, during which time his research was sponsored by the Office of Naval Research for eight years. He has a Ph.D. in industrial engineering and operations research from the University of California at Berkeley. His e-mail address is <averill@simulation.ws> and his website is <www.averill-law.com>.