Computer Modeling & Simulation

Ángel A. Juan Pérez

PID_00152600

Ángel A. Juan Pérez

Angel A. Juan is an Associate Professor of Simulation and Data Analysis in the Computer Science, Multimedia and Telecommunication Studies at the Open University of Catalonia (UOC). He holds a Ph.D. in Applied Computational Mathematics (UNED), an M.S. in Information Systems & Technology (UOC), and an M.S. in Applied Mathematics (University of Valencia). His research interests include both service and industrial applications of Modeling & Simulation, Quantitative Data Analysis, and Computer Supported Mathematical e-Learning. For more information, please visit his site at http://ajuanp.wordpress.com.

Table of Contents

Course overview

Welcome to this graduate course on **Computer Modeling and Simulation** (CMS), a hybrid discipline that combines knowledge and techniques from Operations Research (OR) and Computer Science (CS) (Figure 1). Due to the fast and continuous improvements in computer hardware and software, CMS has become an emergent research area with practical industrial and services applications. Today, most real-world systems are too complex to be modeled and studied by using analytical methods. Instead, numerical methods such as simulation must be employed in order to study the performance of those systems, to gain insight into their internal behavior and to consider alternative ("what-if") scenarios. Applications of CMS are widely spread among different knowledge areas, including the performance analysis of computer and telecommunication systems or the optimization of manufacturing and logistics processes. This course introduces concepts and methods for designing, performing and analyzing experiments conducted using a CMS approach. The course discusses the proper collection and modeling of input data and system randomness, the generation of random variables to emulate the behavior of the real system, the verification and validation of models, and the analysis of the experimental outputs.
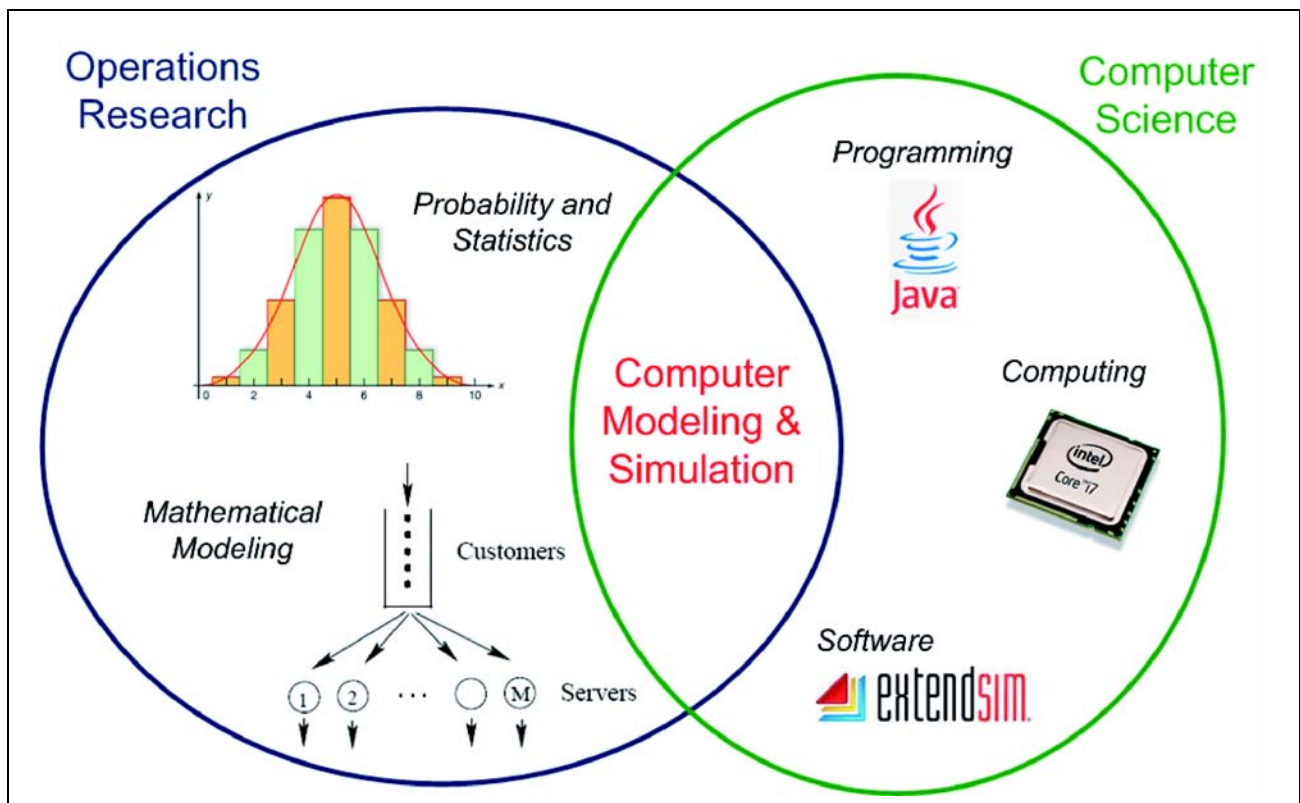
Figure 1. CMS combines Operations Research and Computer Science

Audience and prerequisites

This course is designed for graduate students in any of the following degrees: Computer Science, Telecommunication Engineering, Business Management, Industrial Engineering, Economics, Mathematics or Physics. A solid **mathematical background** is required –it is assumed that the student has successfully completed at least three undergraduate courses on mathematics, one of them in probability and statistics. Also, some programming skills (e.g.: Java or C#) are desirable although not necessary –students will be able to choose between a more programming-focused curricula or a more software-focused curricula. Finally, the student must be able to read technical documentation written in English.

Goals and objectives

The main **goals** of this course are:

- To introduce students in the prolific research field of **computer-supported system modeling and simulation**. In particular, this course offers a set of powerful system modeling and analysis tools (concepts, techniques and skills) that students can use both in their research and professional careers.

- To develop students' **modeling**, **analytical-thinking and synthesis skills**. In particular, throughout the course students will have to: model systems or processes in order to analyze them, read research papers and develop their own scientific articles.

Course **objectives** are derived from the course goals and designed to be assessable. By the end of this course, students should be able to:

- Apply scientific thinking to the analysis of complex systems and processes.

- Comprehend important concepts in computer modeling and simulation.

- Model uncertainty and randomness by means of statistical distributions.

- Form a hypothesis and design a computer experiment to test it.

- Collect and model data, estimate errors in the results and analyze simulation outputs.

- Understand how computers generate (pseudo-)random numbers and variates.

- Realize the application scope and limitations of computer simulation techniques.

- Employ statistical techniques to construct scientific statements and conclusions.

- Construct, verify and validate system and processes models.

- Synthesize in a scientific paper the main ideas and results of a research project.

Learning resources

Probably the two most relevant books in the CMS arena are those from Law (2007) and Banks et al. (2009). In fact, this course is based on the latter book (Figure 2), which contains detailed explanations of most CMS concepts and techniques as well as an abundant collection of application examples. These examples cover fields such as computer and telecommunication networks, manufacturing and logistics processes and system reliability and availability issues.

Figure 2. Basic learning resources for the course

Another fundamental source of CMS learning resources is the **Winter Simulation Conference** or WSC (www.wintersim.org). Every year, this international event is the central meeting point for hundreds of CMS scientists, engineers and users who share their knowledge and research results. Papers published in the WSC proceedings are not only available from the most prestigious databases (ACM Digital Library or IEEE Digital Library) but they can also be freely downloaded from the **INFORMS Simulation Society** website (http://www.informs-sim.org). The WSC proceedings contain top-quality research papers covering almost any imaginable CMS topic –from introduction tutorials to advanced applications– and, therefore, they will constitute an invaluable learning resource for both the beginner and the advanced researcher or practitioner.

Finally, it is worthy to remember that Internet is today a great source of learning materials for almost every knowledge area. In particular, there are thousands of resources on issues related to system modeling and analysis, probability and statistics, programming languages and simulation software, etc., that can be useful for students registered in this course. Two websites that usually offer accurate mathematical-related contents are **Wikipedia** and **MathWorld** (http://mathworld.wolfram.com).

Methodology

Instructors will provide guidance and support to students through the **UOC Virtual Campus** by posting notes on the virtual-class forums, answering students' doubts, submitting recommendations, providing personalized feedback whenever possible and proposing specific lectures, debates and assessment activities as scheduled in the course syllabus.

It is expected that students maintain an **open**, **constant and creative attitude** throughout the course, showing real interest in learning new –and sometimes difficult– concepts, working hard on their learning process according to instructors' recommendations, and developing an original research that might eventually be synthesized in a scientific article.

When performing mathematical and/or statistical operations –which are frequently necessary to solve certain problems and exercises–, use of **mathematical and/or statistical software** is not only allowed but strongly recommended. CMS students, researchers and practitioners should be familiarized with at least one mathematical software (e.g.: Wiris, Maple, Mathcad, Mathematica, Matlab, GNU Octave, etc.) and one statistical software (e.g.: Minitab, Excel, SPSS, R, S-Plus, etc.).

Evaluation system

This is an online graduate course with a demanding educational program based on quantitative contents. Therefore, in order to achieve the proposed academic goals and objectives, students must actively participate in the on-line forums and complete the different activities that constitute the **continuous evaluation process**. This continuous evaluation process will consist of a set of deliverables (typically three or four) that must be submitted to the instructor on or before their associated deadlines. The last deliverable will be an **original scientific paper** developed by the student and supervised by his/her instructor. The paper can be written in English, Spanish or Catalan and must address a specific CMS topic. After a reviewing and improvement process, the best papers could be submitted to international conferences or journals for their possible publication. More details about this continuous evaluation process will be provided in the course syllabus at the beginning of the academic semester.

1. Introduction to Computer Modeling & Simulation
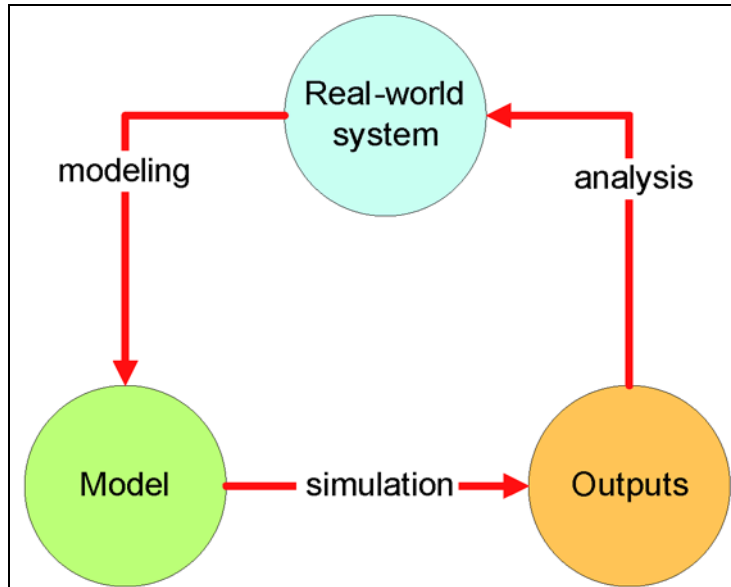
Figure 3. Modeling, Simulation and Analysis

1.1. Book chapters to study

Read carefully Chapter 1, **"Introduction to Simulation"**, from Banks et al. (2009). This chapter describes the main concepts and ideas associated with computer modeling and simulation.

1.2. Summary of basic concepts

- A **system** is defined as a group of objects that are joined together in some regular interaction or interdependence toward the accomplishment of some purpose.

- The **state** of a system is defined to be a collection of variables necessary to describe the system at any time.

- An **event** is defined as an instantaneous occurrence that might change the state of the system.

- A **discrete system** is one in which the state variables change only at a discrete set of points in time.

- The behavior of a real-world system as it evolves over time can usually be studied by developing a computer simulation **model**. This model usually

takes the form of a set of mathematical or logical assumptions concerning the operation of the real system.

- **Simulation** involves the generation of an artificial history of the model representing the system and the observation of that artificial history to draw inferences concerning the operating characteristics of the real system.

- **Discrete-event simulation (DES)** refers to the simulation of systems in which the state variable changes only at a discrete set of points in time.

- Once developed and validated, a model can be used to investigate a wide variety of **"what-if" questions** about the real system: potential changes to the system can first be simulated, in order to predict their impact on system performance.

- The availability of special-purpose simulation languages and software and of massive computing capabilities have made simulation one of the most **widely used** and accepted tools in operations research and systems engineering.

- Most modern systems and processes are so **complex** that their internal behavior can be treated only through simulation.

- Simulation takes **data** on the real system. If no data is available, simulation might not be advised.

- In contrast to optimization models, simulation models are numerically **"*run*"** rather than analytically solved.

- The **applications** of simulation are vast, including fields such as Manufacturing, Computer and Telecommunication Networks, Health Care, Logistics and Transportation, Construction Engineering, Risk Analysis or Business Process. The Winter Simulation Conference (www.wintersim.org) is an excellent way to learn more about the latest in simulation applications and theory. WSC programs with full papers are available from http://www.informs-sim.org/wscpapers.html.

- Some of the typical **stages** in any simulation study are the following ones: data collection, input modeling, experimental design, model implementation, model verification, model execution, model validation, and output analysis.

1.3. Articles to study

Banks, J. (1999): "Introduction to Simulation". In: *Proceedings of the Winter Simulation Conference*, pp. 7-13.

Banks, J. (2001): "The Future of Simulation". In: *Proceedings of the Winter Simulation Conference*.

Goldsman, D. (2007): "Introduction to Simulation". In: *Proceedings of the Winter Simulation Conference*, pp. 26-37.

Ingalls, R. (2002): "Introduction to Simulation". In: *Proceedings of the Winter Simulation Conference*, pp. 7-16.

Raychaudhuri, S. (2008): "Introduction to Monte Carlo Simulation". In: *Proceedings of the Winter Simulation Conference*, pp. 91-100.

Sanchez, P. (2006): "As simple as possible, but no simpler: a gentle introduction to simulation modeling". In: *Proceedings of the Winter Simulation Conference*, pp. 2-10.

Saraph, P. (2004): "Future of Simulation in Biotechnology Industry". In: *Proceedings of the Winter Simulation Conference*, pp. 2052-2054.

Shannon, R. (1998): "Introduction to the Art and Science of Simulation". In: *Proceedings of the Winter Simulation Conference*, pp. 7-14.

1.4. Exercises

1) Get a copy of a recent WSC Proceedings and report on the different applications discussed in an area of interest to you.

2) Using your favorite search engine, search the web for the following terms: "discrete event simulation", "web based simulation", "supply chain simulation" and "parallel simulation". Prepare a brief report explaining each of these terms.

3) Search for other international conferences focused on computer modeling and simulation and prepare a brief report with basic information on them.

1.5. Complementary lectures

Optionally, you can also read Chapters 2 to 4 from Banks et al. (2009). Chapters 2 ("Simulation Examples") and 3 ("General Principles") will help you to acquire a better understanding of the concepts previously introduced as well as to learn more about different real-world CMS applications. Especially interesting is the discussion in Chapter 3 regarding the **event scheduling** and **process interaction** approaches to CMS. Notice that while the former offers a more "centralized" approach (all events are scheduled and managed by a

main procedure), the latter is a more "distributed" one (each simulation object schedules and manages its own events list). Chapter 4, "Simulation Software", offers a general review of simulation software which includes general-purpose languages (e.g.: Java or C#), simulation programming languages (e.g.: GPSS/H) and simulation environments (e.g.: Arena, Extend or Excel @Risk).

Specific examples of use of some simulation programming languages and environments can be found at the WSC proceedings.

2. Probabilistic models
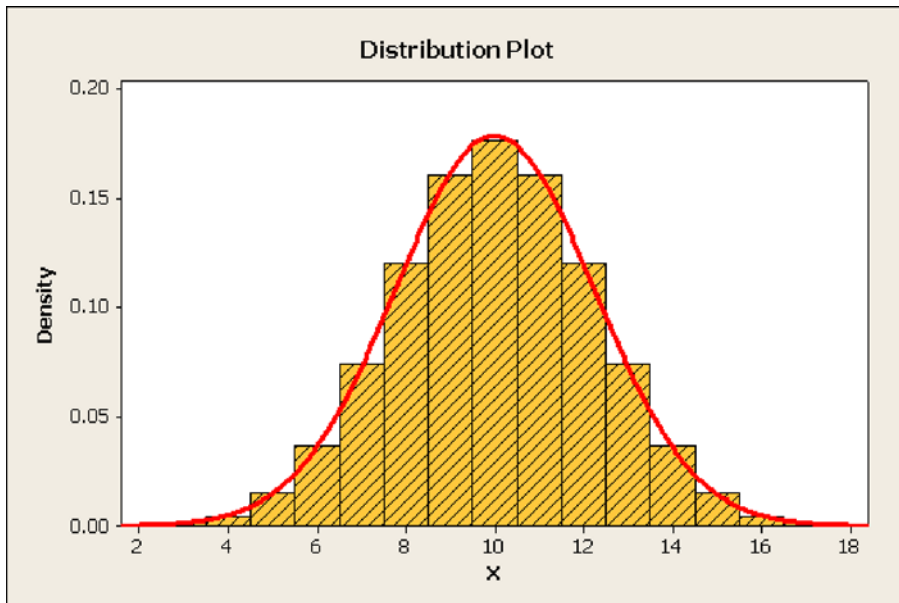
Figure 4. Two probability distributions

2.1. Book chapters to study

Read carefully Chapter 5, **"Statistical Models in Simulation"**, from Banks et al. (2009). This chapter contains a review of basic probability concepts and review the most frequently used statistical distributions in CMS. The chapter also defines the concept of a Poisson process and explains how it is related to an exponential distribution. Finally, the chapter also discusses the use of empirical distributions.

2.2. Summary of basic concepts

- Uncertainty (variation) in real-world systems is usually modeled by means of **random variables**. Each of these random variables follows a statistical distribution.

- For each random variable in the model, data from the real system must be collected and analyzed. Ideally, this data should be fitted by a theoretical **statistical distribution**. If no theoretical distribution can be chosen to adequately represent the input data, then an empirical distribution should be used instead.

- Fitted statistical distributions or probabilistic models of input data are then used in **generating random events** in a simulation experiment.

- A random variable X is **discrete** if the number of possible values of X is finite or countably infinite. A **continuous** random variable is one with an uncountably infinite number of possible values.

- A statistical distribution is univocally characterized by its **cumulative distribution function** (CDF), which usually includes a set of **parameters**. Some frequent parameters of statistical distributions are the mean or expected value, the variance and the mode.

- Some relevant **discrete distributions** are: the Bernoulli distribution, the binomial distribution, the geometric distribution, the negative binomial distribution and the Poisson distribution.

- Some relevant **continuous distributions** are: the uniform distribution, the exponential distribution, the gamma distribution, the Erlang distribution, the normal distribution, the Weibull distribution, the triangular distribution, the lognormal distribution and the beta distribution.

- Consider a queueing system where arrivals are random and independent events that occur one at a time. For a given time t ⩾ 0, N(t) denotes the number of arrivals by time t. Then, the counting process, {N(t), t ⩾ 0}, is said to be a **Poisson process** with mean rate λ.

- The random variable X = N(t) follows a **Poisson distribution** with parameter λ = λt. Also, the random variable Y = "system interarrival times" follows an **exponential distribution** with mean 1/λ.

- The exponential distribution is **memoryless**, meaning that the probability of a future arrival in a time of length s is independent of the time of the last arrival (the probability of the arrival depends only on the length of the time interval, s).

- An **empirical distribution** is a distribution whose parameters are the observed values in a sample data. An empirical distribution may be used when it is impossible to fit a random variable by some particular parametric distribution.

2.3. Articles to study

Barton, R.; Schruben, L. (2001): "Resampling methods for input modeling". In: *Proceedings of the Winter Simulation Conference*, pp. 372-378.

Kelton, D. (2007): "Representing and generating uncertainty effectively". In: *Proceedings of the Winter Simulation Conference*, pp. 38-42.

Law, A. (1997): "ExpertFit: Total support for simulation input modeling". In: *Proceedings of the Winter Simulation Conference*, pp. 668-673.

Law, A. (2003): "How to conduct a successful simulation study". In: *Proceedings of the Winter Simulation Conference*, pp. 66-70.

Law, A.; McComas, M. (2001): "How the ExpertFit distribution-fitting software can make your simulation models more valid". In: *Proceedings of the Winter Simulation Conference*, pp. 256-261.

Tenorio, M.; Nassar, S.; Freitas, P.; Magno, C. (2005): "Recognition of continuous probability models". In: *Proceedings of the Winter Simulation Conference*, pp. 2524-2531.

Vaughan, T. (2008): "In search of the memoryless property". In: *Proceedings of the Winter Simulation Conference*, pp. 2572-2576.

2.4. Exercises

1) From the book Banks et al. (2009), try to complete the following exercises related to chapter 5: 5.1, 5.2, 5.6, 5.9, 5.14, 5.15, 5.19, 5.25, 5.30, 5.36, 5.40, 5.47, 5.50 and 5.51.

2.5. Complementary lectures

From Wikipedia:

- Probability distribution:
  http://en.wikipedia.org/wiki/Probability_distribution

- List of probability distributions:
  http://en.wikipedia.org/wiki/List_of_probability_distributions

From MathWorld:

- Statistical distribution:
  http://mathworld.wolfram.com/StatisticalDistribution.html

3. Queueing models
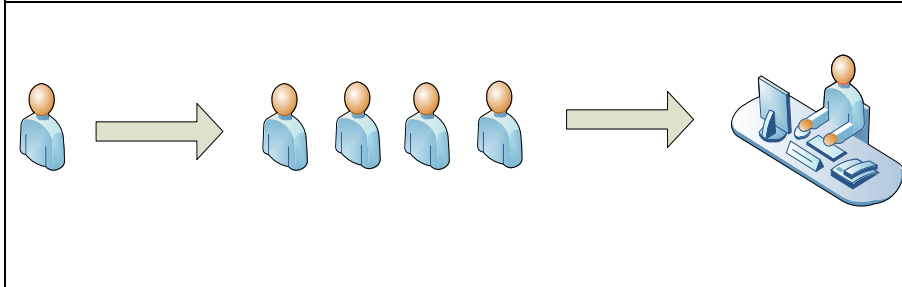

Figure 5. A simple queueing system

3.1. Book chapters to study

Read carefully Chapter 6, **"Queueing Models"**, from Banks et al. (2009). This chapter discusses the general characteristics of queues (waiting lines), the important performance measures, how to use simulation to estimate the mean measures of performance, the effect of varying the input parameters (what-if scenarios), and the mathematical solution of some basic queueing models.

3.2. Summary of basic concepts

- In a typical **queueing model** customers (people, machines, requests, etc.) arrive from time to time and join a queue (waiting line), are eventually attended by a server (person, machine, computer, etc.), and finally leave the system. Queueing models, whether solved mathematically –which is only possible for relatively simple models– or analyzed through simulation, provide the analyst with a powerful tool for designing and evaluating the performance of queueing systems.

- Queueing models have found widespread use in the analysis of service facilities, production and material-handling systems, computer and communications systems, and many other situations where congestion or competition for scarce resources can occur. **CMS techniques** may be used to generate a number of artificial histories of a complex system and to obtain estimates of desired system-performance measures.

- Typical **measures of queueing system performance** include server utilization (percentage of time a server is busy), length of waiting lines, and delays of customers. Quite often the decision maker is involved in tradeoffs between server utilization and customer satisfaction in terms of line lengths and delays.

- Some **parameters that characterize a queueing system** are: the calling population (which might be finite or infinite), the system capacity (either limited or not), the interarrival times between consecutive customers (usually modeled by a probability distribution), customers behavior (they might reject to enter a too long queue or leave after waiting for a while), queue discipline (FIFO, LIFO, random, shortest processing time first, service according to priority, etc.), queueing mechanism (systems with more than one queue are frequent), service times (also modeled by probability distributions), number of servers, service mechanism (e.g.: each customer might have to be attended by more than one server), etc. Notice that a queueing system can be really complex to model and analyze even when using simulation.

- Classical **long-run measures of performance** of queueing systems are: the average number of customers in the system (L) and in the queue (Lq) –notice that "system" includes queue plus service–, the average time spent in the system (W) and in the queue (Wq) per customer, the server utilization (p), the proportion of customers who are delayed in the queue longer than t_0 time units, the proportion of customers turned away because of capacity constraints, the proportion of time the waiting line contains more than k_0 customers, and the total costs of the system –which should include not only direct operational costs but also costs due to customers' dissatisfaction, etc.

- A queueing system is said to be in statistical equilibrium, or **steady state**, if the probability that the system is in a given state is not time dependent. Under some assumptions (exponential interarrival times, FIFO discipline, etc.), it is possible to mathematically solve several steady-state queueing models of infinite population, namely: M/G/1, M/M/c, M/M/c/N. Models using exponentially distributed interarrival times are usually known as **Markovian models**.

3.3. Articles to study

Anderson, E.; Morrice, D. (2002): "Capacity and backlog management in queuing-based supply chains". In: *Proceedings of the Winter Simulation Conference*, pp. 1302-1305.

Klungle, R. (1999): "Simulation of a claims call center: a success and a failure". In: *Proceedings of the Winter Simulation Conference*, pp. 1648-1653.

Kreutzer, W.; Hopkins, J.; Mierlo, M. (1997): "SimJAVA: a framework for modeling queueing networks in Java". In: *Proceedings of the Winter Simulation Conference*, pp. 483-488.

Medeiros, D.; Swenson, E.; DeFlitch, C. (2008): "Improving patient flow in a hospital emergency department". In: *Proceedings of the Winter Simulation Conference*, pp. 1526-1531.

Mehrotra, V.; Fama, J. (2003): "Call center simulation modeling: methods, challenges, and opportunities". In: *Proceedings of the Winter Simulation Conference*, pp. 135-143.

Mesquita, M.; Hernandez, A. (2006): "Discrete-event simulation of queues with spreadsheet: a teaching case". In: *Proceedings of the Winter Simulation Conference*, pp. 2277-2283.

3.4. Exercises

1) From the book Banks et al. (2009), try to complete the following exercises related to chapter 6: 6.2, 6.3, 6.9, 6.24, 6.25 and 6.27.

3.5. Complementary lectures

From Wikipedia:

- Queueing theory:
  http://en.wikipedia.org/wiki/Queueing_theory

4. Random numbers and variates

$$X_{i+1} = \left(aX_i + c \right) \bmod m \qquad i = 0, 1, 2, \ldots$$

$$R_i = \frac{X_i}{m} \qquad i = 1, 2, \ldots$$

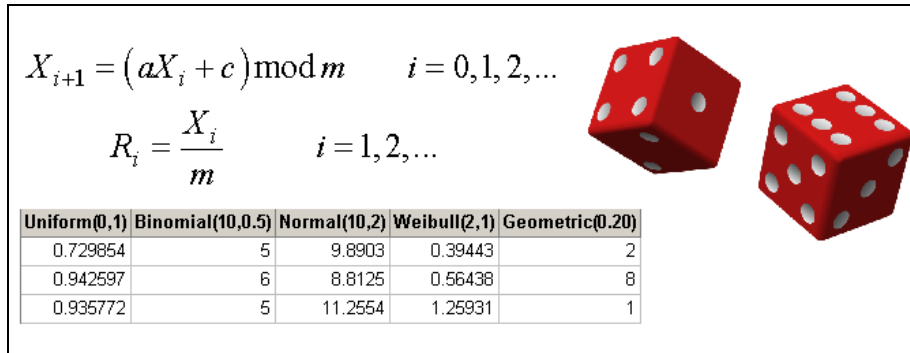| Uniform(0,1)	Binomial(10,0.5)	Normal(10,2)	Weibull(2,1)	Geometric(0.20)
0.729854	5	9.8903	0.39443	2
0.942597	6	8.8125	0.56438	8
0.935772	5	11.2554	1.25931	1

Figure 6. Generation of random numbers and random variates

4.1. Book chapters to study

Read carefully Chapters 7, **"Random-Number Generation"**, and 8, **"Random-Variate Generation"**, from Banks et al. (2009). The former chapter discusses the generation of random numbers and their subsequent testing for randomness. The latter chapter shows how random numbers are used to generate a random variable with a desired probability distribution.

4.2. Summary of basic concepts

- **Random numbers** are a fundamental ingredient in any discrete-event simulation. Most computer and simulation languages generate random numbers that are used to generate event times and other random variables.

- A sequence of random numbers must have two important statistical properties: **uniformity** –i.e., they follow a continuous uniform distribution in the $(0,1)$ interval– and **independence** –i.e., each new number is independent of its predecessors.

- Since they use deterministic algorithms, computers do not generate real random numbers. Computers generate **pseudo-random numbers**, which imitate the ideal properties of uniform distribution and independence as closely as possible.

- The **cycle length** or period of a random number generator (RNG) represents the length of the random number sequence before previous numbers begin to repeat themselves in an earlier order. The generator algorithm should have a sufficiently long cycle.

- The **linear congruential method** produces a sequence of integers, X_1, X_2, ... between 0 and m-1 by following a recursive relationship $X_{i+1} = (aX_i + c)$ mod m (i = 0, 1, 2, ...). Then, random numbers between 0 and 1 can be generated by considering $R_i = X_i / m$ (i = 1, 2, ...). The initial value X_0 is called the **seed**. If c 0, the form is called the **mixed** congruential method; otherwise it is known as the **multiplicative** congruential method. The selection of the values a, c, m, and X_0 drastically affects the statistical properties and the cycle length. Some of these generators can reach periods close to $2{\star}10^9$.

- **Combined linear congruential generators** mix two or more multiplicative congruential generators to increase the cycle length while maintaining good statistical properties. Some of these generators can reach periods close to $3{\star}10^{57}$.

- A long-period RNG can be divided into a set of disjoint sequences or **streams**. That way, a single random-number generator with k streams can act like k distinct virtual random-number generators. This can be useful, for example, when comparing two or more alternative systems via simulation, since it allows assigning specific streams to generate the same random behavior in each of the simulated systems –e.g.: in a queueing system, one stream could be assigned to generate interarrival times, another stream could be assigned to generate service times, etc.

- **Frequency tests** are used to check for uniformity of random numbers. In particular, the Kolmogorov-Smirnov (for both large and small samples) or the chi-square test (only for large samples) can be employed to compare the distribution of the set of numbers generated to a uniform distribution. Also, **autocorrelation tests** can be used to check for independence of random numbers.

- Well-known procedures exist for obtaining **random variates** from a variety of widely-used continuous and discrete distributions. Most of these procedures assume that a source of uniform [0,1] random numbers is readily available.

- The **inverse-transform technique** is a general methodology that can be used to sample from the exponential, the uniform, the Weibull, the triangular distributions, empirical distributions and discrete distributions.

- The **acceptance-rejection technique** is a general methodology that can be used to sample from the uniform, Poisson and gamma distributions among others.

- Some probability distributions –e.g. the Normal, the Lognormal or the Erlang– have special properties that allow for specific-purpose variate-generation techniques.

4.3. Articles to study

Ewald, R.; Rossel, J.; Himmelspach, J.; Uhrmacher, A. (2008): "A plug-in-based architecture for random number generation in simulation systems". In: *Proceedings of the Winter Simulation Conference*, pp. 836-844.

L'Ecuyer, P. (1986): "Efficient and portable 32-bit random variate generators". In: *Proceedings of the Winter Simulation Conference*, pp. 275-277.

L'Ecuyer, P. (1997): "Uniform random number generators: a review". In: *Proceedings of the Winter Simulation Conference*, pp. 127-134.

L'Ecuyer, P. (2001): "Software for uniform random number generation: distinguishing the good and the bad". In: *Proceedings of the Winter Simulation Conference*, pp. 95-105.

L'Ecuyer, P.; Buist, E. (2005): "Simulation in Java with SSJ". In: *Proceedings of the Winter Simulation Conference*, pp. 611-620.

L'Ecuyer, P.; Meliani, L.; Vaucher, J. (2002): "SSJ: A framework for stochastic simulation in Java". In: *Proceedings of the Winter Simulation Conference*, pp. 234-242.

Leydold, J.; Hormann, W. (2006): "Black-box algorithms for sampling from continuous distributions". In: *Proceedings of the Winter Simulation Conference*, pp. 129-135.

4.4. Exercises

1) From the book Banks et al. (2009), try to complete the following exercises related to chapters 7 and 8: 7.4, 7.6, 7.7, 7.8, 7.14, 7.20, 8.1, 8.3, 8.10, 8.17 and 8.19.

4.5. Complementary lectures

- SSJ for simulation in Java:
  http://www.iro.umontreal.ca/~simardr/ssj/indexe.html

From Wikipedia:

- Random numbers:
  http://en.wikipedia.org/wiki/Random_number_generation

- Random variates:
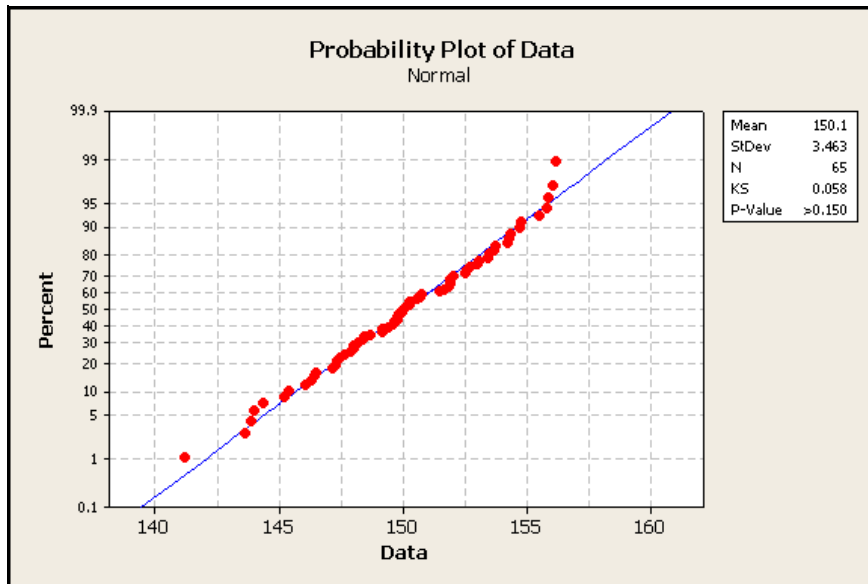 http://en.wikipedia.org/wiki/Random_variate

5. Input data modeling

Figure 7. A Q-Q plot for testing normality of data

5.1. Book chapters to study

Read carefully Chapter 9, **"Input Modeling"**, from Banks et al. (2009). This chapter discusses the fundamental steps that must be performed in order to develop accurate probabilistic models from the real-system data. These models will be used to generate the random events during the simulation process.

5.2. Summary of basic concepts

- In real-world simulation applications, coming up with appropriate **distributions** for input data is a major task. Faulty models of the inputs will lead to outputs whose interpretation could give rise to misleading recommendations ("garbage-in-garbage-out").

- There are **four steps** in the development of a useful model of input data: (a) collect data from the real system of interest, (b) identify a probability distribution to represent the input process, (c) choose parameters that determine a specific instance of the distribution family, and (d) evaluate the chosen distribution and the associated parameters for goodness of fit.

- A frequency distribution or **histogram** is useful in identifying the shape of a distribution. However, a histogram is not as useful as a **Q-Q plot** –which measures the difference between the empirical CDF and the fitted CDF– for evaluating the **fit** of the chosen distribution.

- After a family of distributions has been selected, the next step is to **estimate the parameters** of the distribution. Basically, these estimators are the maximum-likelihood estimators based on the raw data.

- **Goodness-of-fit tests** (e.g.: chi-square, Kolmogorov-Smirnov or Anderson-Darling) provide helpful guidance for evaluating the suitability of a potential input model. Careful must be taken when performing one of these tests: if very little data are available, then a goodness-of-fit test is unlikely to reject any candidate distribution; but if a lot of data are available, then a goodness-of-fit test will likely reject all candidate distributions.

- The **chi-square** goodness-of-fit test formalizes the intuitive idea of comparing the histogram of the data to the shape of the candidate density function. The test is valid for large sample sizes and for both discrete and continuous distributional assumptions when parameters are estimated by maximum likelihood.

- The **Kolmogorov-Smirnov** (KS) test formalizes the idea behind examining a Q-Q plot. It allows testing any continuous distributional assumption for goodness of fit. The KS test is particularly useful when sample sizes are small.

- A test that is similar to the KS is the **Anderson-Darling** (AD) test, which is also based on the difference between the empirical CDF and the fitted CDF.

- Most modern statistical packages implement several goodness-of-fit tests. Usually, these packages provide a **p-value** for the test statistic. This p-value is the significance level at which one would just reject the null hypothesis that the fit is correct.

- Unfortunately, in some situations, a simulation study must be undertaken when there is not possible to obtain data from the system. When this happens, it is particularly important to examine the **sensitivity** of the results to the specific models chosen.

5.3. Articles to study

Barton, R.; Cheng, R.; Chick, S.; Henderson, S.; Law, A.; Leemis, L.; Schemeiser, B.; Schruben, L. (2002): "Panel on current issues in simulation input modeling. In: *Proceedings of the Winter Simulation Conference*, pp. 353-369.

Biller, B.; Nelson, B. (2002): "Answers to the top ten input modeling questions". In: *Proceedings of the Winter Simulation Conference*, pp. 35-40.

Cowdale (2006): "Lessons identified from data collection for model validation". In: *Proceedings of the Winter Simulation Conference*, pp. 1280-1285.

Law, A.; McComas, M. (1997): "ExpertFit: total support for simulation input modeling". In: *Proceedings of the Winter Simulation Conference*, pp. 668-673.

Leemis (2000): "Input modeling". In: *Proceedings of the Winter Simulation Conference*, pp. 17-25.

Leemis (2004): "Building credible input models". In: *Proceedings of the Winter Simulation Conference*, pp. 29-40.

Schmeiser (1999): "Advanced input modeling for simulation experimentation". In: *Proceedings of the Winter Simulation Conference*, pp. 110-115.

5.4. Exercises

1) From the book Banks et al. (2009), try to complete the following exercises related to chapter 9: 9.6, 9.7, 9.10, 9.11, 9.13, 9.14, 9.16, 9.20, 9.21, 9.31.

5.5. Complementary lectures

From Wikipedia:

- Goodness-of-fit:
  http://en.wikipedia.org/wiki/Goodness_of_fit

- Chi-square test:
  http://en.wikipedia.org/wiki/Pearson's_chi-square_test

- Kolmogorov-Smirnov:
  http://en.wikipedia.org/wiki/Kolmogorov-Smirnov

From MathWorld:

- Chi-square test:
  http://mathworld.wolfram.com/Chi-SquaredTest.html

- Kolmogorov-Smirnov:
  http://mathworld.wolfram.com/Kolmogorov-SmirnovTest.html
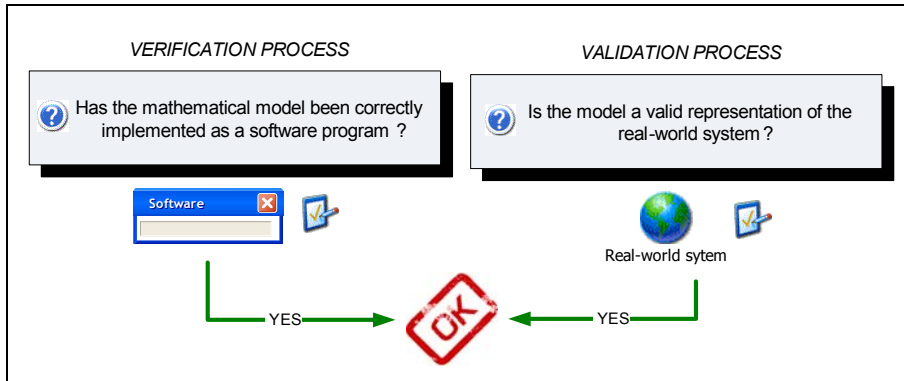
6. Verification and validation

Figure 8. Key questions of the verification and validation processes

6.1. Book chapters to study

Read carefully Chapter 10, **"Verification, Calibration and Validation of Simulation Models"**, from Banks et al. (2009). This chapter describes methods that have been recommended and used in the verification and validation processes, which constitute two fundamental phases of any simulation experiment.

6.2. Summary of basic concepts

- **Verification** is concerned with building the model correctly. It proceeds by comparison of the conceptual or mathematical model to the computer representation that implements that conception. It asks the questions: Is the mathematical model implemented correctly in the simulation software? Are the input parameters and logical structure of the model represented correctly?

- **Validation** is concerned with building the correct model. It attempts to confirm that a model is an accurate representation of the real system. Validation is usually achieved through the calibration of the model, an iterative process of comparing the model to actual system behavior and using the discrepancies between the two, and the insights gained, to improve the model.

- The goal of the validation process is twofold: (1) to produce a model that represents true system behavior closely enough for the model to be used as a **substitute** for the actual system for the purpose of experimenting with the system; and (2) to increase to an acceptable level the **credibility** of the model, so that the model will be used by managers and other decision makers.

- Verification and validation of simulation models are of great importance. Decisions are made on the basis of simulation results; thus, the **accuracy** of these results should be subject to question and investigation.

6.3. Articles to study

Balci, O. (1997): "Verification, validation and accreditation of simulation models". In: *Proceedings of the Winter Simulation Conference*, pp. 135-141.

Carson, J. (2002): "Model verification and validation". In: *Proceedings of the Winter Simulation Conference*, pp. 52-58.

Caughlin, S. (2000): "An integrated approach to verification, validation, and accreditation of models and simulations". In: *Proceedings of the Winter Simulation Conference*, pp. 872-881.

Metz, M.; Jordan, J. (2001): "Verification of object-oriented simulation designs". In: *Proceedings of the Winter Simulation Conference*, pp. 600-603.

Rabe, M.; Spieckermann, S.; Wenzel, S. (2008): "A new procedure model for verification and validation in production and logistics simulation". In: *Proceedings of the Winter Simulation Conference*, pp. 1717-1726.

Robinson, S. (1997): "Simulation model verification and validation: increasing the users' confidence". In: *Proceedings of the Winter Simulation Conference*, pp. 53-59.

Sargent, R. (2007): "Verification and validation of simulation models". In: *Proceedings of the Winter Simulation Conference*, pp. 124-137.

6.4. Exercises

From the book Banks et al. (2009), try to complete the following exercises related to chapter 10: 10.1, 10.2, 10.3 and 10.8.

6.5. Complementary lectures

From Wikipedia:

- Hypothesis testing:
  http://en.wikipedia.org/wiki/Statistical_hypothesis_testing

- Confidence interval:
  http://en.wikipedia.org/wiki/Confidence_interval

From MathWorld:

- Hypothesis testing:
  http://mathworld.wolfram.com/HypothesisTesting.html

- Confidence interval:
  http://mathworld.wolfram.com/ConfidenceInterval.html
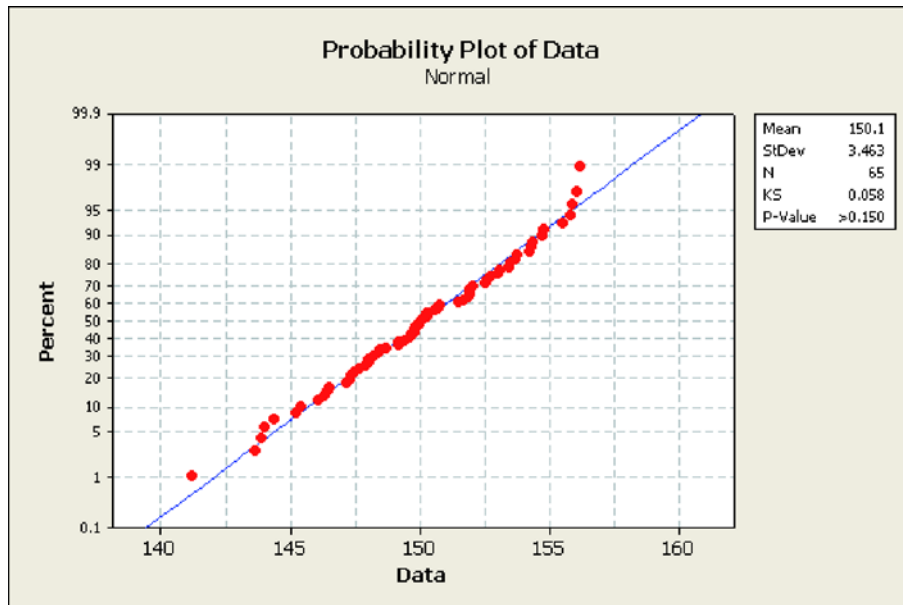
7. Output analysis

Figure 9. Graph of system performance under different scenarios

7.1. Book chapters to study

Read carefully Chapter 11, **"Estimation of Absolute Performance"**, from Banks et al. (2009). This chapter deals with the analysis of a single system from the simulation outputs.

7.2. Summary of basic concepts

- **Output analysis** is the examination of data generated by a simulation. Its purpose is either to predict the performance of a system or to compare the performance of two or more alternative system designs.

- The need of **statistical output analysis** is based on the observation that the output data from a simulation exhibits random variability when random number generators are used to produce the values of the input variables. Therefore, if the performance of the system is measured by a **parameter** θ, the result of a set of simulation replications will be an **estimator** $\hat{\theta}$ of θ. The precision of the estimator can be measured by its standard error or by a confidence interval for the parameter.

- Usually, the analyst wants to study **steady-state**, or long-run, properties of the system –that is, properties that are not influenced by the initial conditions of the model at time 0. In those situations, **"warm-up" periods** –in-

itial periods in which the system has not reach its steady state already–should be identified and removed from the analysis.

- For each parameter of interest, an unbiased **point estimator** should be provided together with its corresponding standard error or, alternatively, by a **confidence interval** for the parameter based on the point estimator.

- It is always possible to reduce the **width of a confidence interval** by increasing the sample size, i.e., the number of replications in the simulation experiment. However, in order to divide by 2 the width of a confidence interval, it is necessary to multiply by 4 the current number of replications.

- Several **variance-reduction techniques** can be employed to reduce the number of replications that are necessary to reach a certain value for the estimation accuracy.

7.3. Articles to study

Alexopoulos, C. (2006): "A comprehensive review of methods for simulation output analysis". In: *Proceedings of the Winter Simulation Conference*, pp. 168-178.

Centeno, M.; Reyes, M. (1998): "So you have your model: what to do next". In: *Proceedings of the Winter Simulation Conference*, pp. 23-29.

Kelton, D. (1997): "Statistical analysis of simulation output". In: *Proceedings of the Winter Simulation Conference*, pp. 23-30.

Kelton, D. (2000): "Experimental design for simulation". In: *Proceedings of the Winter Simulation Conference*, pp. 32-38.

Law, A. (2007): "Statistical analysis of simulation output data: the practical state of the art". In: *Proceedings of the Winter Simulation Conference*, pp. 77-83.

Lecuyer, P. (2006): "Variance reduction in the simulation of call centers". In: *Proceedings of the Winter Simulation Conference*, pp. 604-613.

Nakayama, M. (2006): "Output analysis for simulation". In: *Proceedings of the Winter Simulation Conference*, pp. 36-46.

7.4. Exercises

1) From the book Banks et al. (2009), try to complete the following exercises related to chapter 11: 11.6, 11.7 and 11.10.

7.5. Complementary lectures

From Wikipedia:

- Standard error:
  http://en.wikipedia.org/wiki/Standard_error_(statistics)

- Confidence interval:
  http://en.wikipedia.org/wiki/Confidence_interval

- Variance reduction:
  http://en.wikipedia.org/wiki/Variance_reduction

From MathWorld:

- Confidence interval:
  http://mathworld.wolfram.com/ConfidenceInterval.html
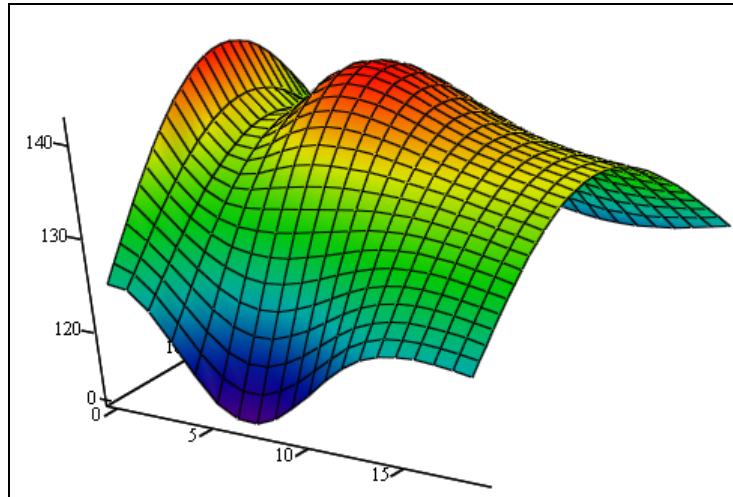
8. Evaluation of alternative designs

Figure 10. A response surface for optimal design

8.1. Book chapters to study

Read carefully Chapter 12, **"Estimation of Relative Performance"**, from Banks et al. (2009). This chapter introduces the comparative evaluation of alternative system designs based on data collected from simulation runs.

8.2. Summary of basic concepts

- One of the most important uses of simulation is the **comparison of alternative system designs**. Because the observations of the response variables contain random variation, statistical analysis is needed to discover whether any observed differences are due to differences design or merely to the random fluctuation inherent in the models.

- A practical strategy to compare two system designs is to compute point and **interval estimates** of the difference in mean performance of a specific system parameter. **Significant differences** between both designs will exist whenever the resulting confidence interval will not contain the zero value.

- It can be sometimes desirable to **reduce the variance** of the estimated difference of the performance measures and thus provide, for a given sample size, more precise estimates of the mean difference that the ones obtained through independent sampling. This variance reduction can be attained by using the **common random numbers** (CRN) technique.

- The basic idea behind the CRN technique is to **assign the same source of randomness** to each random variable in both systems. Each random number used in one model for some purpose should be used for the same purpose in the second model –e.g. by dedicating one random-number stream to a specific purpose. Notice that this implies the necessity of synchronizing the use of random numbers throughout the simulation experiment, which is not a trivial task.

- Comparison of several system designs is also possible by using the Bonferroni approach, **Analysis of Variance** (ANOVA) techniques or even regression analysis.

- In the so called **"optimization through simulation"**, the goal is to optimize (minimize or maximize) some measures of system performance when system performance is evaluated by running a computer simulation. When the inputs that define the behavior of a system or process are not deterministic but stochastic (i.e. they are random variables), the output value representing the system performance, Y, will depend upon the specific values of the inputs and, therefore, it cannot be optimized. In those cases, however, the goal will be to optimize the **expected value**, $E[Y]$, or long-run average of that performance parameter.

- Many **heuristics** and **random-search algorithms** have been developed for deterministic optimization problems. Most of these heuristics use randomness as part of their strategy, e.g.: genetic algorithms (GA), tabu search (TS), simulated annealing (SA), GRASP, ant colony optimization (ACO), etc. These algorithms do not guarantee finding the optimal solution, but they have shown to be very effective on difficult, practical problems, e.g.: vehicle routing problems, scheduling problems, etc.

- There are many additional topics of potential interest in the realm of statistical analysis techniques relevant to simulation. **Experimental design models** –whose purpose is to discover which factors have a significant impact on the performance of system alternatives–, or **variance reduction techniques** –which are methods to improve the statistical efficiency of simulation experiments– are just two of these topics.

8.3. Articles to study

Baesler, F.; Araya, E.; Ramis, F.; Sepulveda, J. (2004): "The use of simulation and design of experiments for productivity improvement in the Sawmill industry". In: *Proceedings of the Winter Simulation Conference*, pp. 1218-1221.

Barra, J.; Ferreira, A.; Leal, F.; Silva, F. (2007): "Application of design of experiments on the simulation of a process in an automotive industry". In: *Proceedings of the Winter Simulation Conference*, pp. 1601-1609.

Carson, Y.; Maria, A. (1997): "Simulation optimization: methods and applications". In: *Proceedings of the Winter Simulation Conference*, pp. 118-126.

Faulin, J.; Gilibert, M.; Juan, A.; Ruiz, R.; Vilajosana, X. (2008): "SR-1: A simulation-based algorithm for the capacitated vehicle routing problem". In: *Proceedings of the Winter Simulation Conference*, pp. 2708-2716.

Juan, A.; Faulin, J.; Marques, J.; Sorroche, M. (2007): "J-SAEDES: A java-based simulation software to improve reliability and availability of computer systems and networks". In: *Proceedings of the Winter Simulation Conference*, pp. 2285-2292.

Kleijnen, J. (2008): "Design of experiments: overview". In: *Proceedings of the Winter Simulation Conference*, pp. 479-488.

Moore, L.; Ray, B. (1999): "Statistical methods for sensitivity and performance analysis in computer experiments". In: *Proceedings of the Winter Simulation Conference*, pp. 486-491.

Olafsson, S.; Kim, J. (2002): "Simulation optimization". In: *Proceedings of the Winter Simulation Conference*, pp. 79-84.

8.4. Exercises

From the book Banks et al. (2009), try to complete the following exercises related to chapter 12: 12.2, 12.3 and 12.22.

8.5. Complementary lectures

From Wikipedia:

- Design of experiments:
  http://en.wikipedia.org/wiki/Design_of_experiments

- Search algorithm:
  http://en.wikipedia.org/wiki/Search_algorithm

From MathWorld:

- ANOVA:http://mathworld.wolfram.com/ANOVA.html

Appendix: Books and Links

8.6. Books

Banks, J.; Carson, J.; Nelson, B.; Nicol, D. (2009): *Discrete-Event System Simulation Prentice Hall*. ISBN: 0136062121. *(This course is based on this book and it constitute an excellent reference for all students and simulation practitioners)*

Faulin, J.; Juan, A.; Martorell, S.; Ramirez-Marquez, E. (eds.) (2010): *Simulation Methods for Reliability and Availability of Complex Systems* (Springer Series in Reliability Engineering). ISBN: 978-1-84882-212-2. *(This book has been co-edited by one of the course's instructors and it shows state-of-the-art applications of simulation techniques to reliability and availability issues)*

Law, A. (2006): *Simulation Modeling and Analysis*. McGraw-Hill Publishing Co. ISBN: 0071255192. *(An outstanding reference for simulation students and researchers)*

McHaney, R. (2009): *Understanding Computer Simulation.* Ventus Publishing ApS. ISBN: 978-87-7681-505-9. *(Free book available at www.bookboon.com)*

8.7. Links

- http://dpcs.uoc.edu
  Website of the DPCS research group (the Academics section contains examples of our CMS-related research lines).

- http://www.wintersim.org/
  Website of the Winter Simulation Conference (WSC).

- http://www.informs-sim.org/
  Website of the Informs Simulation Society (contains papers published at the WSC).

- http://mathworld.wolfram.com/
  Website containing mathematical resources.

- http://cv.uoc.edu/webapps/calculadora/es/index.html
  Wiris mathematical software at the UOC Virtual Campus

- http://www.itl.nist.gov/div898/handbook/index.htm
  Online handbook of Engineering Statistics